Exploitation of NMR in the analysis and screening of fragment ligands for an SH2 domain

by

Jonathan D. Taylor

Department of Biochemistry & Molecular Biology
University College London
Gower Street
London

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy from the University of London

December 2005

UMI Number: U602473

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U602473
Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

A fragment-based approach to drug design has recently emerged in which small chemical groups that bind to adjacent sites on a receptor are identified, optimised, and linked together to generate a high-affinity lead compound. Due to its sensitivity, nuclear magnetic resonance (NMR) is a particularly useful technique for the detection of low-affinity fragment binding. An aim of this project was to explore further the utility of NMR with regards to fragment ligand binding using a well-understood model system – the Src Homology 2 (SH2) domain from *v*-Src kinase. This choice was supported by a large body of structural, functional, and thermodynamic data, coupled with a decade of pharmaceutical investigation into several SH2 domains. Our ongoing biophysical studies of *v*-Src SH2 required determination of the apo solution structure of this domain. A high quality structural ensemble was obtained using standard NMR techniques and a combination of manual and automated assignment methods. The internal hydrogen bonding, ionisation states, and backbone dynamics of the apo *v*-Src SH2 domain was also explored using NMR. The perturbation of *v*-Src SH2 backbone chemical shifts and dynamics by interaction with fragment ligands yielded insights into SH2 domain binding behaviour and specificity. Computational approaches were used to identify potential fragment ligands for *v*-Src SH2. A small library of these molecules were screened *in vitro* using a recently-proposed ^{19}F-NMR competition screening approach, which was optimised for the detection of low-affinity fragments. Follow-up NMR and calorimetry experiments confirmed the screening results and provided further characterisation of the novel fragment ligands. Such compounds may be useful as phosphotyrosine mimetics in SH2-related drug design. A novel ^{31}P-based screening experiment was also proposed. These studies have furthered our understanding of the SH2 domain, in terms of its binding specificity and drug design, and of the NMR screening approaches useful for fragment-based lead discovery.

Acknowledgements

First and foremost, I would like to acknowledege and thank deeply the God of the universe, the designer and sustainer of all life, the author of the laws of nature, the architect of the cell. The more I investigate and consider the biochemistry of the human cell, the more I am awed by its design, its complexity, its precision, and inevitably, by the one who made it. This thesis is not only a piece of scientific research – it is also a treatise on a miniscule part of His massive creation.

There are many people to whom I am indebted for the support and guidance kindly given during the course of this doctorate. I am very thankful to those who have provided overall scientific instruction and vision, particularly my supervisor Professor John Ladbury. Thank you for helping me understand the wider context of my research and for encouraging me to seek my own path. My research was also partly supervised and sponsored by *Celltech* (now part of *UCB*) and I am very grateful to Dr Will Pitt, Mr Philip Gilbert, and Dr Richard Taylor for their enthusiasm and additional support throughout, and in particular the time they gave up during my 3-month placement in Cambridge. This formed an important part of my studies and I'm very grateful to you and to *Celltech* for providing me with that experience and for permitting an almost exclusive focus on the SH2 domain model system. I have always felt very privileged in having a multidisciplinary supervisory team contributing to my research.

I have also learned a great deal from the instruction and assistance of Drs Mark Williams and Matthew Cliff at *UCL*. Thank you for all your support and encouragement for the practical help you gave in performing experiments and analysing results. It has been a challenging three years in biophysics but I have emerged somewhat wiser thanks to your consistent support and interest. In the laboratory I have received the assistance of many past and present Ladbury-group members, in particular Drs Zamal Ahmed, Roger George, and Shusuke Ono. I thank Radwan Fawaz for his help with the interpretation of dynamics data and his relentless enthusiasm for my project. Thanks also to Dr Abdessamad Ababou for providing some support in completing the NMR structure and in the calculation of pK_a values. I would also like to commend the high level of technical support provided by Duncan Mackenzie

and Donovan Binns, the departmental IT staff. Whilst at *Celltech* I benefited from the helpful advice of Dr Colin Groom and Alícia Higueruelo.

Throughout my PhD, and in particular these last few months, I have always received and appreciated the unswerving support of my family and friends. I would like to acknowledge them as important contributors to my work and thank them greatly for listening, for seeking to understand my work, and for providing distraction and perspective. Finally, an immeasurable amount of support to this research has come from my incredible wife, Alison, who did everything she could to encourage and enable me in my research. Although many of the tasks involved have seemed alien and unfathomable, she has consistently believed in me and motivated me to work to the best of my ability. Thank you, Ali, for your patience, strategic guidance, prayers, and delicious food. At many times, progress has seemed Sisyphean, but you kept me grounded in life and faith; it would have been ten times harder without you.

Jonathan D. Taylor
December 2005

Table of contents

List of Figures

List of Abbreviations

ACD	Available Chemical Database
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical Shift Anisotropy
DCL	Dynamic Combinatorial Library
DiFMUP	6,8-difluoro-4-methyl-umbelliferylphosphate
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ELISA	Enzyme-linked immuno-sorbent assay
FAXS	Fluorine chemical shift anisotropy and exchange for screening
FDA	Food and Drug Administration
FID	Free Induction Decay
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
HSQC	Heteronuclear Single Quantum Coherence
HTS	High Throughput Screening
ITC	Isothermal Titration Calorimetry
MES	2-(4-Morpholino)ethanesulfonic acid
MUP	4-methyl-umbelliferylphosphate
MW	Molecular Weight
NCE	New Chemical Entity
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PDB	Protein Data Bank
PDGF	Platelet-Derived Growth Factor
PFG	Pulse-field gradient
ppm	Parts-per-million
PTP	Protein Tyrosine Phosphatase
RMSD	Root Mean Squared Deviation
SAR	Structure-activity relationship
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SGI	Silicon Graphics Inc.
SH2	Src homology 2
SH3	Src homology 3
SPA	Scintillation Proximity Assay
SPR	Surface Plasmon Resonance
STD	Saturation Transfer Difference
TINS	Target immobilised NMR screening
TRIS	2-Amino-2-(hydroxymethyl)-1,3-propanediol
TROSY	Transverse Relaxation-Optimized Spectroscopy

Chapter 1

Introduction

This thesis explores the use of nuclear magnetic resonance as a screening technique suitable for fragment-based drug development. The work forms part of a continuing collaboration between *Celltech* (now part of *UCB*) and the *Molecular Biophysics* group at *UCL*, the broad aim of which is to deepen our understanding of protein-ligand interactions. These studies made use of a suitable model system, a Src homology 2 (SH2) domain, which has been a significant focus within our research group over a number of years.

Celltech are an international pharmaceutical company that has become interested in this novel, fragment-based approach to the discovery and development of therapeutic compounds. Central to the application of this method is a suitable ligand screening technique that is capable of detecting the inherently weak interactions. Nuclear magnetic resonance (NMR) spectroscopy is particularly suited to this task and avoids many of the pitfalls associated with current high-throughput approaches. As a consequence, the pharmaceutical industry has extended the traditional use of NMR into the field of compound screening through extensive conceptual, experimental, and technological development during the previous decade. More recently, other biophysical techniques, such as x-ray crystallography and mass spectrometry, have been applied as alternative fragment screening methods. *Celltech* therefore wish to explore the applicability of NMR in the analysis of weak-binding fragments, in the development of suitable compound libraries, and in the interface between computational techniques and *in vitro* follow-up screening. This thesis demonstrates how NMR can be used to probe ligand binding sites by detecting changes in either protein- or ligand-based spectra and it also provides further characterisation of the model system. Chapters 3 and 4 report on research, performed at *UCL*, into SH2 domain structure, mobility, ionisation states, and responses to ligand binding. Chapters 5 and 6 describe computational and NMR-based screening studies carried out at *Celltech* (Cambridge, UK).

This introduction describes four subjects relevant to this project. Firstly, an overview of the current approaches to, and difficulties experienced in therapeutic ligand discovery is provided. The fragment-based method is then described fully in terms of theoretical validity and practicality, and its various merits and drawbacks discussed. A comparison of the various fragment screening techniques leads into the second part, which provides a fuller description of experimental approaches in NMR-based compound screening. The third section introduces the model system in terms of its role in cell signalling and human disease, its structural and molecular biology, and finally the possibilities for therapeutic intervention. The chapter concludes with an overview of SH2 domain inhibitors, which entered pharmaceutical development about 13 years ago. These efforts have been hindered by a marked intolerance towards phosphotyrosine replacement groups that are required to improve the oral bioavailability and stability of such compounds. Moreover, due to the presence of many ($\sim$115)[1] homologous SH2 domains within the human cell, coupled with a general lack of understanding of signal transduction, it has proved difficult to achieve a high level of compound specificity[2].

1.1 Therapeutic ligand discovery

Prior to the final decade of the twentieth century, the principal approach to identifying therapeutic compounds was by screening for ability to induce a desired phenotypic change. Whilst this approach was successful in many cases, increased levels of drug quality control has caused many candidate therapeutics to be rejected during development, either due to unfavourable pharmacokinetic properties, or unacceptable side effects that may reflect poor selectivity of the compound. In order to reduce this high attrition rate of compounds during (pre-)clinical development* many pharmaceutical companies have adopted a target-directed† approach which seeks to characterise disease at the molecular level using molecular and structural biology, genomic and proteomic profiling, cell biology, physiology, bioinformatics and *in vivo* studies. Target identification and validation has become a research field in its own right and has benefited from recent advances in various '-omics' technologies and, particularly, the elucidation of the human DNA sequence. Thorough characterisation of a therapeutic target is advisable for at least two major reasons. Firstly, a robust causal link between modulation of a particular molecule and the recovery of a healthy phenotype can be established, ensuring compound efficacy. Secondly, knowl-

*Pre-clinical development refers to the analysis of the compound efficacy, toxicology and metabolism using isolated cells or tissues, or using animal models of disease. Clinical development is the stage prior to drug approval in which safety and efficacy is investigated using pool of human volunteers. This consists of at least three levels of clinical trials which require several years to complete.

†A pharmaceutical 'target' is the molecule with which the intended drug will interact.

edge of the potential or actual side effects caused by modulating certain molecular activities can prevent future wastage of resources on projects that are likely to fail. However, it should be noted that the final approved drug will itself have a separate range of side effects that cannot be addressed at this stage.

Once a particular therapeutic target has been validated, the process of lead-discovery aims to identify compounds that interact with this molecule and bring about the desired change in biochemistry. Over the last 10-20 years, this has involved high-throughput screening (HTS) of large compound collections using rapid, assay-based experiments and follow-up structural analysis. This phase in the history of drug discovery has been (and to some extent still is) dominated by the belief that lead compounds can be obtained by screening as many compounds as possible. This has resulted in considerable investment in combinatorial chemistry and HTS equipment, as well as compound libraries themselves. The choice of compounds to be screened also continues to be a highly active research field as workers have sought to optimise the rate of lead discovery. Central to this approach are the concepts of 'drug-likeness' and compound diversity. The former notion is used to filter out compounds that are unlikely to result in acceptable therapeutics, as defined by those currently approved. Whilst such an approach would reduce wastage, it can restrict innovation, miss useful compounds and perhaps lead to patent infringement issues. The opposing idea of maximal compound diversity aims to explore novel areas of 'chemical space'. Both concepts remain subjective, and are variably defined in the literature. Computer-aided drug design methods have also progressed significantly both in their sophistication and usage throughout the industry. Perhaps the two most widely-used techniques are structure comparisons (pharmacophore and chemical similarity searching) and molecular docking. Many other applications facilitate the curation and selection of library compounds and analyse the vast quantities of data generated in a typical HTS program.

Acceptable lead compounds are subjected to further optimisation and pre-clinical development. Structural analysis of the interaction with the target is frequently used to suggest modifications that can be assessed in binding assays. Compounds closely related to identified leads can also be screened for activities. Iterative improvements in affinity, potency, selectivity and pharmacokinetic properties result in disclosure of a drug candidate and preparation for clinical evalulation[‡]. Compounds approved for human clinical trials are subjected to a series of investigations into matters of safety (Phase I) and efficacy (Phase II), before being tested in a large cohort of patients (Phase III). Although approval of a drug for marketing requires successful completion of Phase III clinical trials, its long-term effectiveness is often explored

[‡]Potency is defined as the pharmacological activity of the compound. Pharmacokinetics refers to the uptake, distribution, metabolism and excretion of the compound.

and compared to that of competitors in Phase IV trials. The United States *Food and Drug Administration* (*FDA*) policy suggests that companies investigate and report on long term adverse side effects. However, a recent editorial in Nature magazine reported that about 70 % of these 1200 studies have yet to begin, suggesting that the *FDA* lacks the political power or will to insist on such studies[3, 4].§

In conclusion, the pharmaceutical industry has, over the past century, utilised a diverse array of approaches towards the development of safe and effective therapeutic ligands for a wide range of disease phenotypes.

1.1.1 Inadequacies in current lead discovery

It is a well-documented fact that despite extensive investment in both compound libraries and HTS equipment, there has not been a concomitant increase in the number of new chemical entities (NCEs)¶ or of drug approvals, which has remained at around 30 per year over the last decade (original source: *FDA* website)[5, 6]. There has been much comment in the literature concerning the failure of large-scale, high-throughput techniques to revolutionise drug design[7, 8, 9]. Although combinatorial chemistry has made a formidable contribution to chemical synthesis research, it has not been properly applied to the field of drug design[7]. The immaturity of the technology coupled with a lack of understanding of which compounds are worth screening has led to the failure of many projects to generate a suitable lead compound[9].

HTS approaches have tended to result in lead compounds that are drug-like in terms of their size and pharmacokinetic properties, however of relatively low potency that precludes entry into pre-clinical development[6]. The lead optimisation process often causes molecular weight increases that reduce drug-likeness to unacceptable levels. Harren Jhoti of *Astex Technology* estimates that 70 % of HTS hits usually fail to progress into lead optimisation due to this reasoning[10]. The shift in focus from screening drug-like compounds to smaller 'lead-like' compounds has only occurred relatively recently[6].

Although combinatorial chemistry has enabled companies to develop large proprietary libraries for screening, it fails to take advantage of the potential chemical diversity space. For molecules consisting of up to 30 non-hydrogen atoms, it has been

§*Merck's* pain killer *Vioxx* serves as an example of the need for effective monitoring of side effects. The drug was approved by the *FDA* in 1999, but two years later agency epidemiologists identified that patients receiving high doses were susceptible to heart attacks and sudden death. Despite this finding, the *FDA* continued to approve *Vioxx* for further clinical indications until *Merck* performed their own study in 2004 and discovered the link with heart attacks and strokes and withdrew the compound from the market. It has been estimated that ~25,000 people suffered early deaths as a result of this oversight. This has led to a large media and political debate concerning the safety of current therapeutics, as well as the unethical dominance of pharmaceutical companies over regulatory bodies.

¶An NCE is a drug that contains no previously approved active moiety

estimated that there are over 10^{60} possible structures[11, 12]. Even the largest current proprietry libraries only contain in the region of 1-2 million compounds. Hann and coworkers demonstrated that increasing the complexity of ligands is likely to reduce the probability of identifying a screening hit[9]. Protein–ligand interactions are determined by a complex balance of many factors – enthalpic and entropic – that together generate the overall free energy of binding. Molecular recognition depends on the features of the ligand being compatible with those of the binding site. Hann *et al.* used a simple model of complexity to show that for a given binding site, the success rate of identifying a ligand is strongly dependent on the complexity of the initial ligand. Complex ligands may contain moieties that confer favourable binding in isolation, however the presence of other groups within the molecule may inhibit such interactions. Thus the probability of identifying a ligand is therefore higher when considering relatively simple compounds.

The extensive investment in HTS technologies has resulted in considerable financial commitments within many pharmaceutical companies towards the maintenance and operation of equipment and the curation of compound libraries. In an effort to reduce screening costs, it has become common to utilise computational pre-screening filters to remove unlikely ligands. Whilst docking, compound similarity, and pharmacophore searching are undoubtedly useful techniques an excessive reliance on computer-driven screening of lead- or drug-like compounds is currently inadvisable.

Thus the current high throughput approaches to lead discovery have not led to the expected increase in pharmaceutical output. This is probably due to several inter-related reasons including conceptual and technological immaturity and an inherently low probability of identifying complex ligands by a mass screening approach. Despite advances in target identification and validation, chemistry and compound screening, the process of lead discovery remains a slow and risky affair. It has been estimated that only 8 % of clinical development candidates reach the market and of these, only one in three yields a return on investment[8]. As the cost of drug development continues to approach the billion dollar mark, some argue that drug discovery is becoming increasingly unsustainable[8].

1.1.2 Fragment approach to drug design

Over the last decade, an alternative approach to drug discovery and design – termed 'fragment-based' – has been proposed in order to break the deadlock encountered with combinatorial chemistry and HTS. In its simplest form, fragment-based methods involve the identification of very small molecules that have affinity for the target of interest, and to optimise, and link them to generate a lead-like ligand of acceptable binding affinity.

A fragment-based approach to drug design is tractable for at least two reasons. Firstly, protein surfaces generally feature a number of distinct interaction subsites, which can be exploited by small molecular fragments. Such cavities differ in size, shape, hydrophobicity, charge and hydrogen-bonding potential. Secondly, a high-affinity compound can be obtained through conjugation of two low-affinity fragment ligands. This idea was formulated in terms of theoretical Gibbs[13] free energy by Page and Jencks over two decades ago[14, 15]. Consider a high-affinity ligand, A—B, which can be divided into two fragments A and B. The binding of each fragment to the receptor can be described in terms of an 'intrinsic binding energy', ΔG_i, and an unfavourable term, ΔG_r, which reflects the loss of rotational and translational entropy (Figure 1.1). For each ligand, the total free energy change on binding is the sum of these two contributions. Although the intrinsic affinity of A—B is more or less additive ($\Delta G_i^{A-B} \approx \Delta G_i^A + \Delta G_i^B$), the respective ΔG_r terms are only weakly determined by molecular weight ($\Delta G_r^{A-B} \approx \Delta G_r^A \approx \Delta G_r^B$) and therefore the total free energy change for binding ligand A—B is much more negative than for the fragments alone, which corresponds to a higher affinity. This model assumes that fragments A and B are linked in such a manner that they able to bind in their optimal configurations without introduction of molecular strain and that the linker itself provides no new favourable interactions.

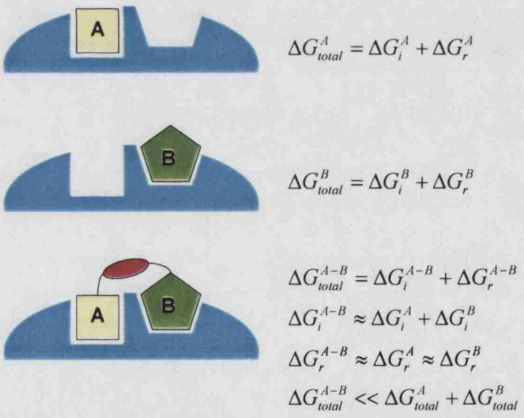


Figure 1.1 Free energy decreases can be achieved through fragment linking.

In principle, Page & Jencks' theory is supported by empirical affinity measurements for A, B and A–B. For example, when developing inhibitors for stromelysin, Hajduk *et al.* identified two fragments with individual K_i values of 17 mM and 20 μM, which, when linked together generated a ligand of K_i = 25 nM (Figure 1.2)[16]. A similar result was obtained for avidin, in which a 0.41 pM inhibitor was generated (Figure 1.2)[17].

Recently, Murray & Verdonk published a further development of Page & Jencks' original theorum[6]. Using the same nomenclature, the binding affinity of molecule

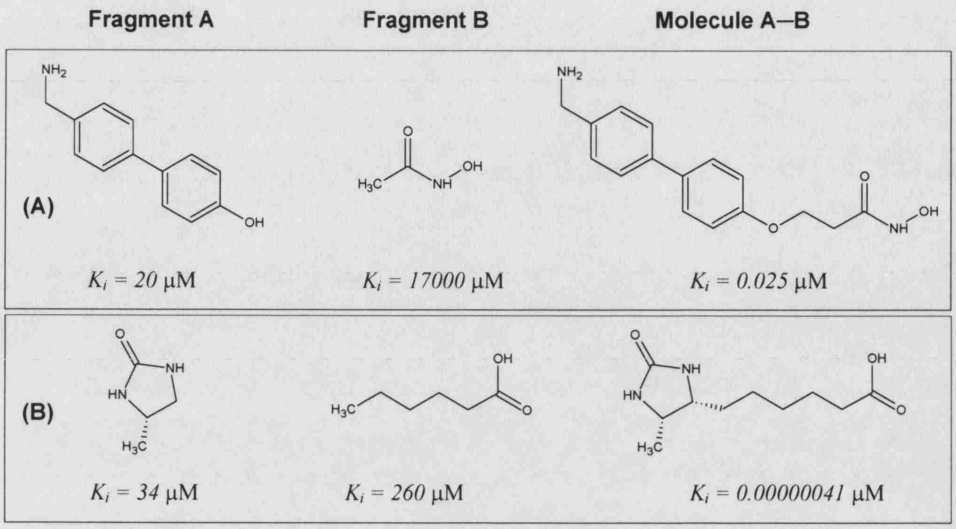


Figure 1.2 Examples of improvements in free energy obtained through fragment linking. Information from Table 1 in Murray & Verdonk[6] was used to generate this figure.

A—B is defined as follows:

$$\Delta G^{A-B}_{total} = \Delta G^{A}_{i} + \Delta G^{B}_{i} + \Delta G_{r} + \Delta G^{A-B}_{rot} + \Delta G^{A-B}_{str} + \Delta G^{A-B}_{binding} \tag{1.1}$$

In this formulation, ΔG_r is a free energy term associated with the loss of rigid body entropy on binding to the target. It is assumed that this is equal for A, B and A—B. Introduction of new rotatable bonds in A—B incurs an entropic penalty (ΔG^{A-B}_{rot}). Free energy is also lost if fragments A or B are not presented into their respective pockets in the lowest energy conformation (ΔG^{A-B}_{str}). Finally, $\Delta G^{A-B}_{binding}$ accounts for any favourable or unfavourable interactions between the protein and the linker.

Murray & Verdonk used the results of previous fragment binding and structural studies to estimate ΔG_r. They ignored situations in which ligand strain and/or linker interactions appeared to affect the overall free energy change on binding (i. e. $\Delta G^{A-B}_{str} \equiv \Delta G^{A-B}_{binding} \equiv 0$) and assumed that $\Delta G^{A-B}_{rot} \approx 2.2$ kJ/mol per bond. Equation 1.1 can be re-written using the observed free energy changes for fragments A and B, and linked compound A—B:

$$\Delta G^{A-B}_{total} = \Delta G^{A}_{total} + \Delta G^{B}_{total} - \Delta G_{r} + \Delta G^{A-B}_{rot} \tag{1.2}$$

The values obtained for ΔG_r ranged between 7 and 29 kJ/mol, which corresponds to a range of 1.5 to 5 orders of magnitude in affinity. Due to the assumptions involved, the authors recommended that ΔG_r be considered in the region of 15-20 kJ/mol, but remember that the value will vary depending on the type of interaction.

In conclusion, Murray & Verdonk suggest that fragments with dissociation constants in the millimolar range are likely to display strongly favourable intrinsic affinities (ΔG_i is large and negative)[6]. Conjugation to another millimolar-affinity fragment using an appropriate linker that delivers each component in an optimal configuration ought to produce a ligand of high affinity, as required for an acceptable lead compound in current pharmaceutical drug design processes.

1.1.3 Fragment linkage and optimisation strategies

Conversion of fragments to high affinity leads is possible, in theory, through provision of the correct linker or other favourable modifications. This section describes several complementary approaches that seek to facilitate this in practice. Throughout, structural information concerning the protein in complex with bound fragments is extremely useful for the efficient direction of fragment elaboration. Knowledge of binding thermodynamics also enables detailed understanding of the energetic effect of particular substitutions. Rees *et al.* (*Astex Technology*) categorize fragment-based approaches to lead discovery and optimisation into four classes that describe the evolution, linking, self-assembly and optimisation of fragments[18]:

Fragment evolution

In this method, a fragment ligand is steadily enlarged by addition of side chains, or other fragments not identified from screening. Modifications that result in affinity improvements become the starting point for further evolution of the molecule. This method has been used to develop inhibitors for several targets, such as Erm methyl transferase and Urokinase (Figure 1.3). It is plausible that fragment evolution could be used to design a side chain that will eventually be used as a linker to fuse two fragments.

Fragment linking

The productive linking of fragments identified from screening presents a significant challenge in the fragment-based ligand design paradigm. Linker design is best considered in terms of binding free energy: An ideal linker will preserve the lowest energy binding conformation of each fragment, with minimal increases in configurational entropy and maximal increases in enthalpy. In this regard, a small, inflexible scaffold that makes favourable interactions with the protein surface is superior to a large, flexible linker. A minimalist approach to linker design is ensures that the characteristics of the ligand, such as molecular weight or polar surface area, do not exceed the upper limits imposed by the pharmaceutical industry.

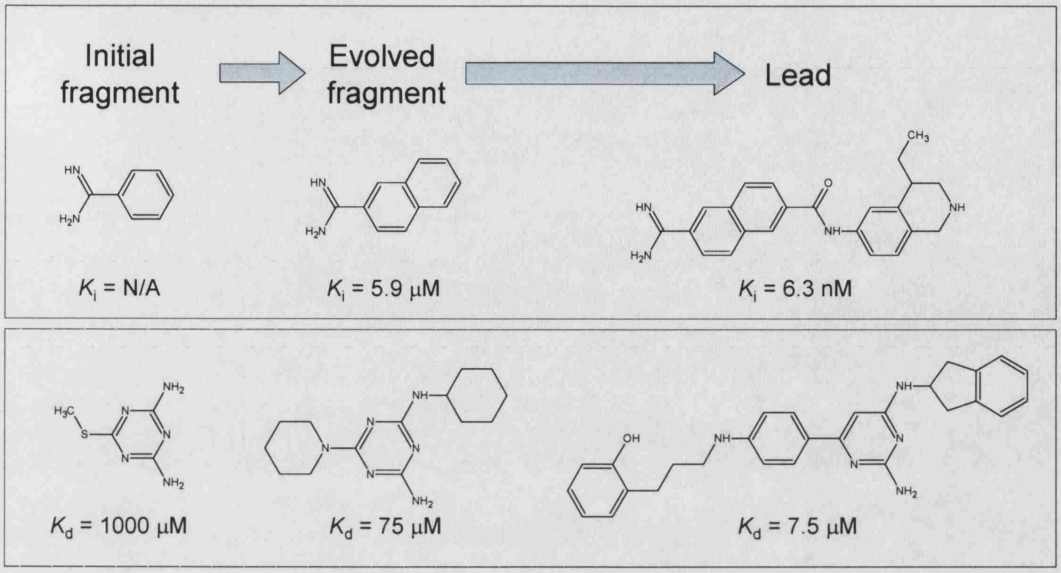


Figure 1.3 Examples of lead identification through fragment evolution. *Top* | Urokinase inhibitor[19]; *Bottom* | Erm methyl transferase inhibitor[20].

Knowledge of the protein structure and binding conformations of fragment ligands is highly instructive in linker design. The exact definition of atoms, including possible contributions from solvent, facilitates modelling of putative linked fragments, and provides a rational basis for maximising linker-protein interactions. A focussed set of compounds can then be synthesised and analysed for their binding ability. Water molecules are found to play an important role in some receptor-ligand interactions[21, 22, 23]. In free energy terms, the entropic cost of trapping a water molecule can be exceeded by an enthalpic gain. This has implications in the design of linkers: attachment of a small group capable of hydrogen bonding directly to the protein surface can displace a trapped water molecule, and may lead to a more favourable free energy.

There are many examples in the literature of lead compounds that are derived from the linked fragment approach. Figure 1.4 indicates two cases in which inhibitors were disclosed for protein tyrosine phosphatase (PTP) 1B[24, 25].

Fragment self-assembly

One of the foundations of modern biology is that molecular interactions are governed by steric and energetic compatibility between surfaces. The properties of the binding site in question thus dictate the ability of compounds to bind. Current drug design efforts, including structure- and fragment-based approaches, require iterative cycles of modification and screening in order to improve binding affinity to acceptable levels. This is because our understanding of molecular interactions in terms of

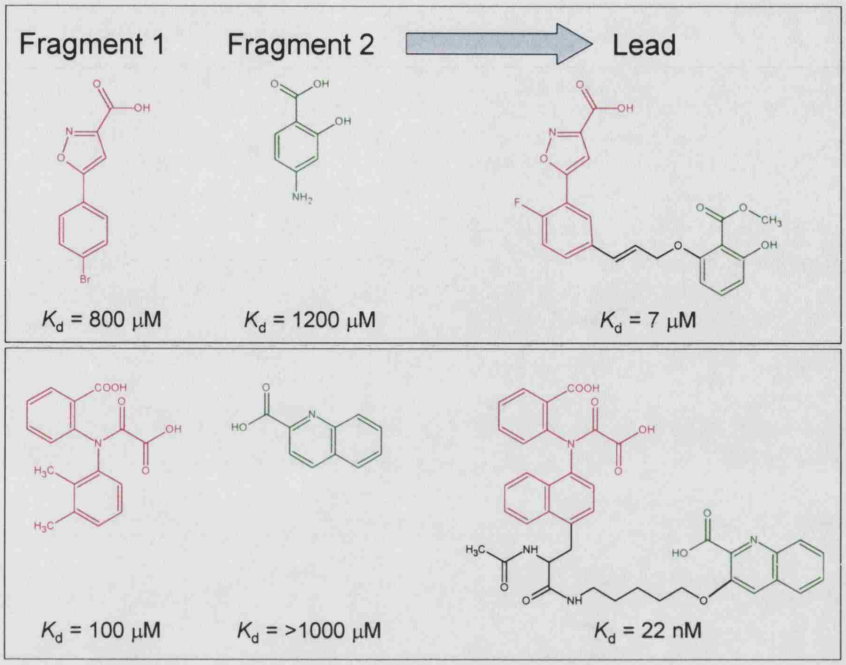


Figure 1.4 Examples of fragment linking in PTP1B inhibitor development.

structure, thermodynamics, solvent effects, mobility are not sophisticated enough to enable routine *de novo* ligand design.

Recently, strategies have emerged in which the prescriptive nature of a given protein binding site is used to guide the assembly of fragments into high affinity ligands[26, 27]. Whereas standard combinatorial chemistry is based on large libraries of pre-synthesised molecules, *dynamic* combinatorial chemistry uses reversible connections between individual components to give access to every possible combination (Figure 1.5). In the presence of a protein target, the continuous exchanging of library components causes the highest affinity fragment combination(s) to be preferentially formed. Exchange between library components is then quenched and the active compound identified.

There are many suitable reversible linking reactions available, such as imine, hemiketal and thiol formation (Figure 1.6)[28]. In an ideal situation, the dynamic combinatorial library (DCL) should be isoenergetic – in the absence of protein, fragment pairing should not be biased by energetically favourable combinations. This can be achieved in practice by separating the fragment from the crosslinking site by a short linker. The equilibrium reaction conditions must not alter the native protein binding site, nor react with the protein itself, as these will lead to incorrect outcomes. Finally, a simple method of quenching the exchange of library components must be available, for example, altering the pH or temperature.

Ramström & Lehn demonstrated the DCL concept by identifying ligands for

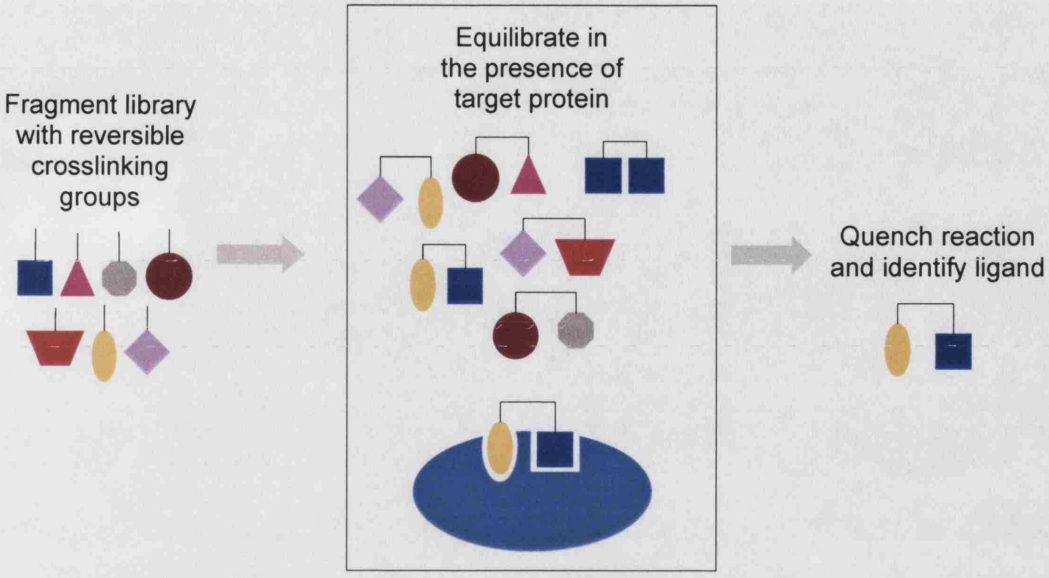


Figure 1.5 Target-guided ligand assembly. Fragments are attached to a cross-linking group that enables random pair-wise formation under certain conditions. Addition of a protein binding site removes high-affinity ligands from the equilibrium, which leads to amplification of the combination within the mixture. Exchange between fragments is quenched, and the active ligand identified.

concanalvalin A[29]. The natural ligand for this protein is a trimannoside moiety, in which two of the rings make important contacts with the protein, whilst the third functions as a linker, and as such can be replaced. Thus, carbohydrate monomers – mannose, galactose, glucose, arabinose and xylose – were attached to a short linker, which terminated in a thiol group. Equilibration of the fragment library at pH 7.4 led to randomised pairing of fragments *via* reversible disulphide bond formation. The linked fragments could be separated and identified using reverse-phase high-performance liquid chromatography (RP-HPLC). Equilibration was continued in the presence of concanavalin-linked sepharose beads, and exchange between disulphides was quenched by acidifying the solution to pH 4. Elution of non-binding compounds from the beads revealed that mannose-containing heterodimers had some binding affinity for concanavalin. Through separation of the protein from the ligands, it was found that the mannose homodimer – closely homologous to the natural ligand – displayed the highest affinity.

Similar schemes have yielded novel leads targeted towards carbonic anhydrase[30], neuraminidase[31], acetycholinesterase[32], and caspase-3[33]. However, the DCL approach is particularly sensitive to the linker design. One can imagine a situation whereby optimal binding fragments exist in the DCL, yet because of inappropriate linkage geometry, they do not form a viable conjugated ligand. It appears to be more sensible to identify individual fragment ligands directed towards adjacent pro-

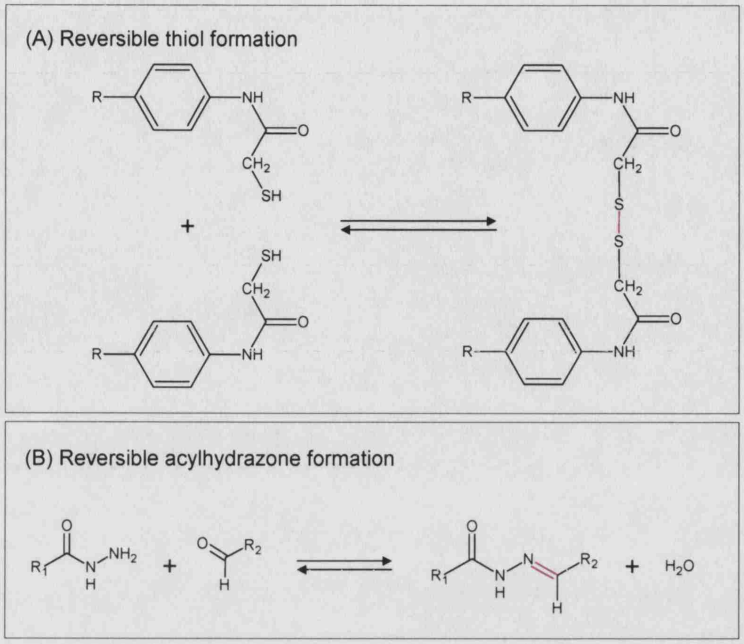


Figure 1.6 Two reversible reactions that can be used to crosslink the attached fragments (denoted R). Note that in example (A), any combination of fragment pairs can be formed, whereas in the second case (B), only R_1—R_2 combinations are possible.

tein subsites first, and then use the DCL approach to determine the optimal linkage strategy by creating a library of fragments connected to a range of different-sized linkers. Indeed, none of the above examples began with completely novel fragments. Instead, existing lead compounds were fragmented and the DCL approach was used to optimise the presentation of fragments into the relevant protein subsites.

Fragment optimisation

This class of fragment-based approach is used to optimise the drug-like properties of a lead candidate other than affinity. Moieties that are associated with poor bioavailability, pharmacokinetics or toxicology can be replaced in a modular fashion in order to prepare a lead compound for clinical trials. For example, a benzamidine fragment was identified in a screen against the S1 pocket of Factor Xa[34]. Using structure-based virtual screening and three rounds of chemical synthesis, a compound with IC_{50} = 16 nM was selected for lead optimisation. The benzamidine fragment was replaced with an analogous group as it is often found to cause poor oral bioavailability. The resultant drug candidate, LY517717, was submitted to Phase II clinical trials.

1.1.4 Patented fragment-based methods

Intellectual property is a highly prized commodity within the pharmaceutical industry. The fact that several companies have sought to develop technologies that facilitate fragment-based methods and claim proprietary rights suggests that the approach is becoming increasingly mainstream. Among companies such as *Plexxikon Inc.*, *Abbott Laboratories*, *Astex Technology*, *Structural GenomiX*, and *Vertex* there has been a large drive towards providing robust fragment-based drug discovery 'platforms' that can revitalise their 'product pipelines' and generate additional income through licensing. Whilst the hype surrounding HTS methods is gradually disappearing, it is being replaced, perhaps prematurely, by enthusiasm towards fragment-based methods. For example, *Plexxikon Inc.* claim that using their *Scaffold-based Drug Discovery*™ procedure, pre-clinical candidate compounds can be synthesised in under six months (source: www.plexxikon.com). Whilst the process has been defined and demonstrated in the literature[35], the success of fragment-based drug design methods can only be adequately assessed during the next 5-10 years as compounds generated using this approach are subjected to clinical trials. Judging by the current level of enthusiasm, it appears that fragment-based enabling technologies will continue to be developed, patented, licensed, acquired and out-sourced in spite of the fact that no drugs derived from fragments have actually reached the stage of governmental approval.[‖]

1.1.5 Advantages of fragment-based drug design

It was demonstrated above why the fragment-based paradigm is valid in theory and how the key element of fragment linking can be achieved in practice. In comparison with current HTS approaches to lead discovery, it is highly advantageous to adopt such a strategy for reasons outlined in the following section. The central concept of the fragment-based method is efficiency – both in the path to a lead compound, and ultimately its quality in terms of pharmaceutical potential. HTS approaches tend to emphasise potency, however hit compounds have often not been suitable for further optimisation. Figure 1.7 shows an approximate empirical relationship between molecular mass and potency for a several compound classes. Hits derived from HTS methods are larger and more potent than fragments, however as a high proportion of atoms within a fragment are involved in binding, the latter group of compounds can be considered more efficient binders[18, 36]. Consequently, when fragments are linked or enlarged to achieve the desired potency, the molecular weight of the compound remains within the acceptable range for leads and/or drug candidates. Improving

[‖]As of August 2004, only one fragment-derived drug candidate (LY517717) has reached Phase II clinical trials[18].

the potency of relatively large HTS hits whilst maintaining or reducing molecular weight has been a challenging task for the pharmaceutical industry, resulting in high attrition rates for screening hits.

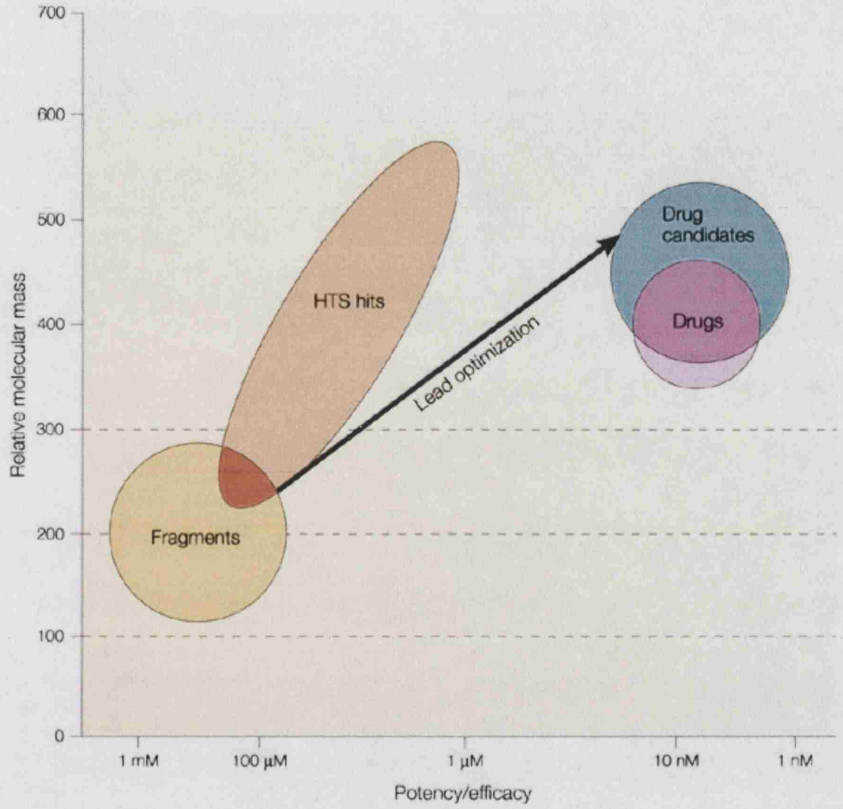


Figure 1.7 Schematic comparison of compound mass and potency ranges for fragments, HTS hits and drugs in clinical development. Figure used with permission from Nature Reviews Drug Discovery[18].

Fragment-based drug design is also advantageous since the number of possible compounds increases exponentially with molecular size[12]. Thus, the diversity of chemical space can be explored more efficiently using smaller molecules. Although combinatorial chemistry is able to perform reactions on a large number of compounds simultaneously, these are often routine in nature, and do not lead to a highly diverse library. Conversely, by investing time in the development of a smaller library of fragments, one can engineer diversity more easily.

As described above, Hann's conceptual model showed that the probability of identifying a ligand for a binding site decreases in an exponential manner as the complexity of the ligand increases. Fragments are therefore more likely to bind to a receptor, albeit weakly, than larger and more complex molecules. In other words, small compounds that form favourable interactions with a protein site are less likely to contain groups that are incompatible with these interactions.

Another advantage of fragment-based approaches over current HTS methods is particularly apparent when considering the practicality and extent of screening required. For instance, in order to analyse every two- or three-way combination of N fragments, one must synthesise and screen N^2 or N^3 molecules (see Figure 1.8). However, by screening individual components for binding activity, and linking fragments that display affinity, far fewer compounds need to be synthesised, saving both time and resources. In the simple example given in Figure 1.8, almost 99.9 % of the pair-wise combinations do not need to be made.

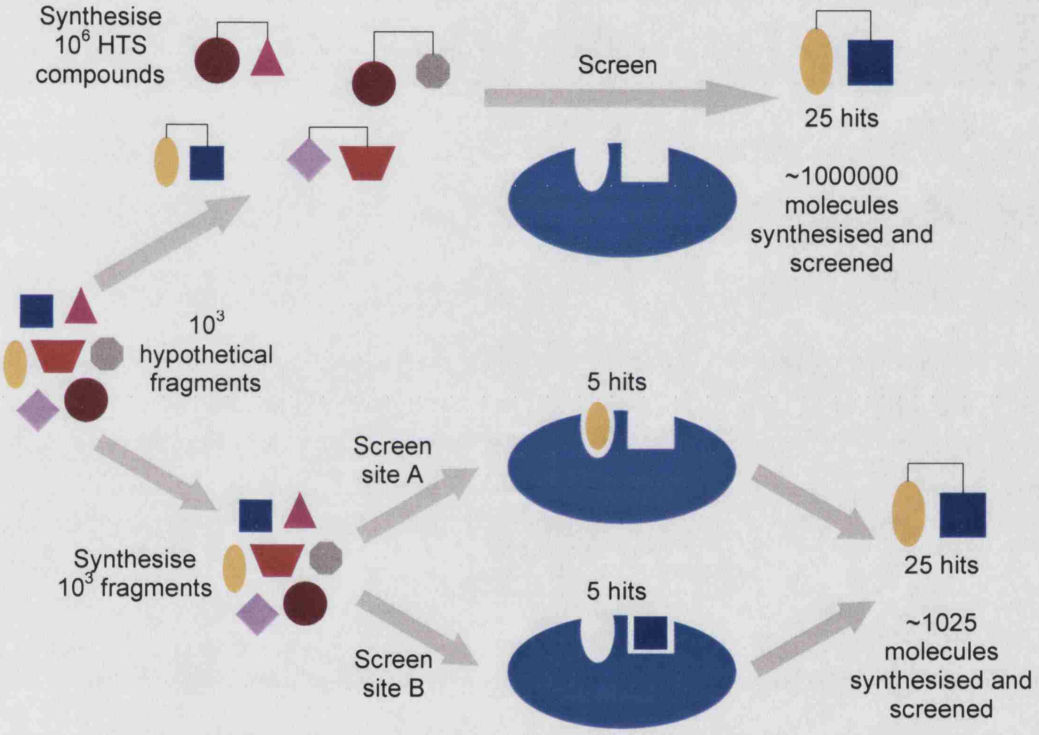


Figure 1.8 Practical advantage of fragment-based drug design. Pairwise conjugation of a 1000-member fragment library generates one million compounds for HTS. Conversely by pre-screening each fragment and linking the resultant hits, far fewer compounds need to be synthesized and screened. Figure adapted from Erlanson & Hansen[37].

The costs involved with acquisition and curation of a small library (<20,000) of fragments are also substantially lower than those associated with million-compound libraries. Since fragments are designed to contain a higher proportion atoms and groups that have the potential to interact with protein targets than HTS compounds, they also tend to be more soluble in aqueous media. The success of compound docking software is expected to decrease with conformational flexibility of the protein and ligand[38]. Thus smaller, more rigid, fragment molecules present a more favourable docking situation compared with larger compounds. Finally, the smaller capital

requirements for compound acquisition, storage and screening for fragment-based methods in drug design enables academic groups to participate in the discovery of suitable lead molecules, even if they lack the resources to fund future clinical trials.

1.1.6 Disadvantages of fragment-based drug design

A major obstacle in the application of the fragment approach the weak binding that is usually associated with very small molecules. Most high-throughput screens utilise a functional or inhibition assay, whereby ligand binding is linked to a colourimetric or fluorimetric result. In such a scheme, weak-binding ligands ($K_d \approx$ mM) fail to elicit a positive response unless they are tested in relatively high concentrations. This leads to solubility problems, both for the compounds and protein target. High concentrations of certain compounds can also directly interfere with colourimetric assays, leading to false-positive results.

Until 1996, the challenge of identifying fragment ligands in a robust manner was too great, and the recent investment in high-throughput technologies distracted the industry. However, a publication from *Abbott Laboratories* described how nuclear magnetic resonance (NMR) could be used to detect fragment binding, and through a rational, structure-based process, build high-affinity ligands[39]. This important development has led to an increased interest in fragment-based approaches to the discovery novel lead compounds and at present, a variety of techniques exist for the selection and screening of fragments, which will be described in turn below.

Rees *et al.* have suggested that fragment-based approaches are design-intensive whilst HTS is resource-intensive[18]. In other words, the former approach is more efficient and more likely to succeed in theory, but it requires a much higher level of scientific direction and skill to achieve. Finally, it should be noted that fragment-based methods are designed to complement and perhaps supersede HTS as the routine method of lead discovery within the pharmaceutical industry. The immediate output from a fragment-based approach is not a drug candidate – lead compounds will still need to be optimised with regards to potency, pharmacokinetics and toxicology prior to clinical trials.

1.1.7 Building fragment libraries

The first stage in implementing a fragment-based strategy is the development of suitable compound libraries. Generally, a 'top-down' approach has been employed whereby the structures of marketed drugs are split into fragments, and the frequency of each particular fragment studied. For instance, Bemis & Murcko of *Vertex Pharmaceuticals* defined atoms within each drug according to one of several sub-types

– rings systems, linkers, frameworks, and side chains (see Figure 1.9)[40, 41]. It was found that 32 of the 1179 unique frameworks were present 50 % of the 5120 drugs analysed. When bond order and atom type were considered, 24 % of drugs contained one of 41 common frameworks.

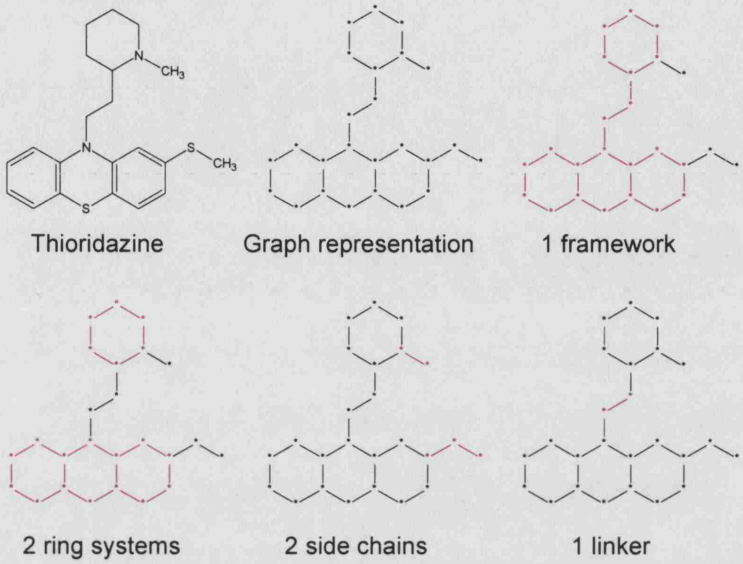


Figure 1.9 Fragmentation of marketed oral drugs into *ring systems*, *linkers*, *side chains*, and *frameworks*. Atom type and bond order are ignored, with each molecule viewed in as a 2D graph. Linkers are atoms that connect two ring systems, side chains are any non-ring, non-linker atom, and frameworks are the union of ring systems and linkers within the molecule. The fragmentation of Thioridazine is provided as an example of each sub-type (coloured red).

Clearly, most drugs can be represented by a fairly small population of molecular structures, and hence these fragments ought to be used to build suitable screening libraries. In this regard, *Vertex* subsequently developed a diverse library of small fragments – termed the SHAPES library – suitable for NMR screening[42]. Currently the library consists of around 200 fragments, which enables rapid screening against low quantities of protein. Combinatorial libraries containing SHAPES fragments typically display HTS hit rates** approximately ten times that of generic libraries[43]. Even higher rates are achieved by designing combinatorial libraries on the basis of SHAPES hits[43]. Since its introduction in 1999, Vertex have used their fragment library to identify inhibitors for a range of protein targets associated with diverse disease phenotypes[42, 44, 45].

In a similar manner, researchers at *Glaxo Wellcome* developed a 'retrosynthetic combinatorial analysis procedure' (RECAP)[46] and Hajduk *et al.* identified frequently-occurring fragments and analysed their *in vitro* propensity for interac-

**hit rate is defined as the percentage of ligands identified from the total number of compounds screened

	'rule of five'	'rule of three'
MW (Da)	≤ 500	≤ 300
Hydrogen bond donors	≤ 5	≤ 3
Hydrogen bond acceptors	≤ 10	≤ 3
cLogP	≤ 5	≤ 3
Polar surface area Å^2	≤ 140	≤ 60
Rotatable bonds	≤ 7	≤ 3

Table 1.1 Comparing Lipinski's 'rule of five' (and subsequent enhancements) with the fragment-orientated 'rule of three' selection criteria. MW corresponds to the molecular weight; cLogP refers to the calculated logarithm of the ratio of solubility in octane and water

tion with proteins[47]. The design of fragment libraries remains an active field of research as some researchers question whether such database analyses actually highlight chemists' preference for well-developed laboratory reactions, rather than suggest intrinsic drug-like properties[48].

In 1997, Lipinski *et al.* released guidelines that defined a range of suitable physico-chemical[††] properties for maximising drug solubility and permeability[49]. This 'rule of five' (Table 1.1) has been widely adopted in the pharmaceutical industry for filtering compounds prior to screening in order to remove molecules that are not 'drug-like'. Others extended the criteria to include the limitation of polar surface area and also the number of rotatable bond terms[50]. Recently, Congreve *et al.* modified Lipinski's scheme in order to make it relevant to fragment selection[51]. The physico-chemical properties of 'successful fragments' – that is, small molecules shown to interact with protein targets – were statistically analysed, and used to establish a 'rule of three' (Table 1.1). Finally, Fattori states that fragments should contain functional groups to enable chemical expansion and optimisation, however these should not be reactive towards the target else false results will be obtained[48].

1.1.8 Fragment screening techniques

There are several methods currently available for screening fragments for binding or inhibitory activity; these will be compared and contrasted. Each technique has inherent strengths and weaknesses, and the choice between them will depend on many factors, such as the availability of equipment, the solubility of compounds, or knowledge of the protein structure.

[††]'Physico-chemical' refers to the structure and chemical properties of a compound.

1.1.8.1 Functional assays

A functional screen seeks to identify ligands that prevent the particular protein of interest from performing its *in vivo* function. Such inhibitors may act in a directly competitive manner by blocking access to the relevant binding site, or it may inactivate the protein through indirect allosteric effects. By screening for ligands that bring about the desired effect, non-inhibitory binders, are ignored. A second advantage of functional screening is that it does not rely on prior knowledge of the structure or mechanism of a target, nor information from other lead compounds. However, at the stage of linking fragments, it is useful to have an understanding of the nature and location of the binding sites.

Existing high-throughput functional screens are not suitable for detection of weak-binding fragments, unless their concentration can be increased to millimolar levels. This often leads to solubility problems for the compounds and assay components, and can also interfere with detection equipment. However, some attempts at screening small fragment libraries have been successful. For example, the Ellman group screened a set of 305 fragments at a concentration of 1 mM against the kinase domain of *c*-Src using a commercially-available ELISA kit[52]. Fragments displaying favourable inhibitory properties (n=37) were used to form a range of heterodimeric compounds, and screened again for activity. One inhibitor displayed an IC_{50} value of 64 nM – an increase of three orders of magnitude over the individual fragments – and was found to be selective towards *c*-Src kinase when assayed with the closely related enzymes Fyn, Lyn and Lck.

Despite this success, fragment-orientated functional screens are rarely reported. Direct detection of binding methods, described below, are providing attractive alternatives to this error-prone approach.

1.1.8.2 NMR-based screening

In the past, application of NMR spectroscopy was limited to the characterisation of molecular structure, dynamics and interactions. For pharmaceutical companies, NMR continues to be an essential tool for identifying reaction products. However, the suggestion and proof that NMR could also be used for screening weak-binding compounds has led to a shift in the perception of NMR, both within the corporate and academic communities, and has provided inspiration for mass-spectrometrists and crystallographers to develop analogous screening protocols.

Although the sensitivity of NMR as a spectroscopic technique is low, it is capable of detecting intermolecular interactions between weak-binding fragments and a macromolecular target. Binding is observed to affect the NMR spectrum of both

species, and there are several NMR experiments available that seek to detect these changes. The array of experiments also enables quantification of binding affinity, location of binding sites, structural analysis of the bound conformation and even full structure determination of the complex. As yet, no single experiment imparts all of this information.

In theory, all NMR spectroscopic parameters can provide a basis for detection of ligand binding, however in practice only those that can be measured in a rapid and sensitive manner are useful. NMR screening approaches can be classified according to whether they detect changes in ligand or receptor resonances. The majority of experiments fall into the former category, exploiting changes in ligand signal relaxation rates, diffusion coefficients and nuclear Overhauser effects (NOEs[‡‡]). The single target-based approach utilises changes in chemical shifts to detect, locate and quantify an interaction. These experiments are described and contrasted in Section 1.2.2–1.2.4 below.

NMR allows direct detection of protein-ligand interactions, hence there is no requirement for complex (and expensive) multi-component biological assays. However, recent developments at *Pharmacia* have resulted in the first NMR screening experiment to monitor the inhibition of an enzyme, through measurement of substrate and product concentrations[53, 54]. The introduction of competition-style experiments, discussed below, has also enabled the identification of inhibitory ligands rather than undesirable inactive binders.

Efficient NMR screening requires a high field spectrometer, with magnetic field of at least than 9.4 T, equivalent to a ^{1}H frequency of 400 MHz. Sample requirements and data collection times are inversely related to field strength, and thus investment in state-of-the-art spectrometers is beneficial. Supplementary robotic and NMR probe technology is also developing rapidly in order to increase the overall screening throughput. Some experiments require prior knowledge of protein or ligand chemical shifts, as well as structural information. Protein concentrations vary between 10 and 300 μM, whilst ligands are typically screened between 50 and 1000 μM. Isotopic labelling of the protein is required for the target-based screening experiment.

Despite its many advantages, NMR screening is not without limitations. As will be described below, many ligand-based experiments are not capable of identifying high-affinity compounds ($K_d \leq 50$ nM). The target-based approach requires that the molecular weight of the protein be less than 40 kDa, although with the appropriate technology, pulse sequences and isotopic labelling strategies, this can be increased to around 100 kDa. This presents a problem for many drug targets, as they are significantly larger than this threshold.

[‡‡]These NMR observables are described in Section 1.2.1

The use of NMR in compound screening has increased steadily and there are now several large pharmaceutical companies, such as *Abbott*, *Vertex*, *Pfizer*, *Novartis*, *Hoffmann-La Roche* and *Triad Therapeutics*, that incorporate fragment-based NMR screening into their drug discovery strategy.

1.1.8.3 Crystallography-based screening

Suggestions that crystallography could be of use in fragment screening were initially criticised because of the twin difficulties of low throughput and the detection of weak-binding compounds[55]. The former problem is being addressed through advances in robotics, crystal generation, computational resources and x-ray technology, whilst the latter issue was dismissed by Mattos & Ringe, who demonstrated that interactions between proteins and organic solvents could be detected by x-ray crystallography[56, 57]. Two companies (*Structural GenomiX* and *Syrrx*), whilst seeking to provide the pharmaceutical industry with structural information concerning a large number of targets, pioneered novel technology for reduction in the time required to synthesise, purify, crystallise and collect x-ray scattering data for a given protein[58]. Thus, inhibitors identified from (non-crystallographic) screening can be used to generate a co-crystal structure in less than a week[58].

In parallel, *Abbott Laboratories* modified their NMR-based screening method for x-ray crystallography, and demonstrated its viability for compound screening[59]. In this pivotal study Nienaber *et al.* soaked urokinase crystals in a solution containing a mixture of fragment compounds. X-ray diffraction data were then collected and used to solve the structure of the protein. As the identity and location of any bound fragment is discernable in the electron density map, library deconvolution is not necessary. A hit was successfully found and optimised, resulting in a potential lead compound with a K_i of 0.37 μM.

Astex have focussed on increasing the throughput of this fragment-based x-ray screening approach by developing several enabling technologies. For instance, their *AutoSolve*™ software can generate protein-ligand complexes from x-ray data in a matter of minutes. With such developments, it appears that crystallography-based fragment screening will become increasingly important to any structure-based drug design strategy. Although compound throughput remains slow relative to functional or NMR-based screening methods, the ease with which structural detail is obtainable may ensure that any ligands identified will automatically be used to generate co-crystal structures for analysis.

A fundamental requirement for crystallographic screening is successful protein crystallisation. Currently, this is problematic for membraneous proteins, such as G-protein coupled receptors (GPCRs), which are the target of 50 % of currently

marketed drugs[60]. Despite this short-coming, there are many proteins of interest to the pharmaceutical industry that are soluble and amenable to crystallisation. At this early stage in development, some difficulties have been encountered with the generation of co-crystals of acceptable quality and with maintaining the integrity of the crystal in the presence of fragment mixtures[61]. Thus complete analysis of a group of ligands was in some cases unachievable.

Another major disadvantage of crystallography-based screening is that the affinity of fragments cannot be precisely measured. Instead proponents of this technique rely on functional screening to validate fragments. Whilst this is straightforward using linked or elaborated fragments of high affinity, this is a problem for weaker-binding, small molecules and may prevent clear understanding of structure-activity relationships (SAR)* during initial stages of the fragment-based approach.

1.1.8.4 Mass spectrometry-based screening

Mass spectrometry (MS) is a highly accurate and sensitive technique for quantification of molecular weight. By careful ionisation and desolvation processes, protein-ligand complexes can be detected, and this provides an opportunity for screening. Initially, researchers at *Ibis Therapeutics* characterised complexes formed between RNA and small molecules, and were able to determine the binding affinity and stoichiometry[62]. Similar studies using protein targets have developed the technique although it is not yet ready for high-throughput compound screening[63, 64].

Erlanson and coworkers at *Sunesis Pharmaceuticals* have utilised an MS-based strategy called 'Tethering'[65]. Potential ligands contain a reactive thiol group, which, when bound to the desired site on the protein, leads to covalent attachment *via* a nearby cysteine. MS is used to detect the increase in molecular weight, and x-ray crystallography facilitates structural characterisation. The tethering approach has also been extended to aid in the identification of fragments that bind to an adjacent site, enabling formation of potent inhibitors[33]. The tethering strategy requires location of a cysteine close to the binding site, which may need to be engineered using site-directed mutagenesis. Additional surface-orientated cysteines should also be removed, or protected, to ensure that only ligands directed towards the target binding site become covalently linked. Screening the thiolated compounds in mixtures will also result in some degree of cross-reactivity, albeit reversible.

*The affinities of closely-related compounds are compared in a typical SAR approach in order to determine which moieties provide maximal affinity.

1.1.8.5 Comparison of screening techniques

Current methods of fragment screening are under continuous development and innovation. Their relative strengths and weaknesses are summarised in Table 1.2. Functional screens represent the original but least sophisticated and least used method of assessing fragment binding. This is largely due to solubility problems experienced at the requisite concentrations. NMR provided many new avenues through which fragment binding could be assessed, revitalising the field. In the past 3-5 years, x-ray crystallography technology developed sufficiently to permit fragment screening and is becoming the favoured option within industry. This is mainly due to the rapid availability of high-resolution structural information concerning ligand binding, although the avoidance of mixture deconvolution and independence on molecular weight are also advantageous features. Others have developed mass spectrometry into a fragment-screening technique, however structural analysis is very often performed using crystallography.

	Functional	**Nuclear Magnetic Resonance**		**X-ray**	**Mass Spectrometry**	
	Screening	*ligand-based*	*target-based*	**crystallography**	*non-covalent*	*tethering*
Protein requirements	Low	Low	High	High	Medium	Medium
Instrumentation (cost)	Normal instruments (low)	High-field NMR (high)	High-field NMR (high)	X-ray source (high)	Mass spectrometer (high)	Mass spectrometer (high)
Throughput	High	Medium	Medium	Low-medium	Medium	Medium
False-positive rate	High	Low-medium	Very low	Very low	Low-medium	Low
Requires knowledge of protein structure?	No	No	Some	Yes	No	Partial
Requires knowledge of protein function?	Yes	No	No	No	No	No
Molecular weight limit?	No	Usually, no	Ideally < 40 kDa but up to 100 kDa is possible	No	Usually <100 kDa	Usually <100 kDa
Information on binding behaviour	Inhibitory potency	Affinity	Binding site and affinity	Structural, but no affinity data	Affinity	Affinity
Deconvolution of mixtures required?	Yes	Variable	Yes	No	No	No
Library modifications required?	No	Sometimes	No	No	No	Yes

Table 1.2 Comparison of fragment screening techniques. This table is adapted from that of Erlanson, McDowell & O'Brien[12].

The three modern approaches to fragment screening – NMR, crystallography, and MS – necessitate acquisition or modification of existing biophysical equipment, or increasingly, out-sourcing to a collaborator. Technological improvements continue to reduce sample requirements and screening timescales for each method. There is a approximate positive correlation between the required financial, temporal and material investments and the amount of useful information gleaned. The strengths of each technique can be combined into various hybrid approaches. For instance, fragment ligands could be identified and validated using NMR and analysed structurally by crystallography. Once affinity has been optimised to micromolar levels,

functional screening can be used to determine inhibitory potential.

Finally, the nature of the target itself is a strong determinant in the choice of screening technique – macromolecules that are difficult to crystallise are excluded from the x-ray method and proteins larger than ~40 kDa inhibit target-based NMR screening.

1.2 NMR-based fragment screening techniques

This section describes in further detail the NMR experimental approaches towards fragment screening. Methods are usually split into two classes, termed *ligand-based* and *target-based*, according to whether the experiment monitors ligand or receptor resonances, respectively. There are currently three types of ligand-based procedures, distinguished according to which NMR observable is followed: signal relaxation rate, translational diffusion rate, and ligand-protein proximity (NOEs). These parameters are described briefly in the following section. Competition equilibria have been employed in order to circumvent problems in the detection of high-affinity compounds within the ligand-based scheme. The target-based approach utilises changes in the nuclear Larmor frequency or chemical shift of receptor resonances on addition of ligand. The concepts behind each experimental technique will be outlined along with consideration of associated strengths and weaknesses. Several reviews have been written describing the use of NMR in fragment screening; see references[44, 66, 67, 68, 69].

1.2.1 Overview of NMR spectroscopy

Elementary particles such as electrons, neutrons, protons and photons possess an intrinsic property called *spin*, characterised by a *spin quantum number* and *spin angular momentum*. The latter parameter is defined by a vector, the direction of which indicates the axis of rotation. In the absence of an external magnetic field, individual vectors randomly sample all possible directions. If a magnetic field is applied then the vectors respond by moving around the macroscopic field axis. Thus the vector traces the path of a cone, maintaining a fixed angle to the field, a phenomenon termed *precession*. The frequency of precession ω^0 (in Hz) is dependent on both the nuclear gyromagnetic ratio γ and the bulk magnetic field strength, B^0:

$$\omega^0 = \frac{-\gamma B^0}{2\pi} \tag{1.3}$$

The proportionality constant γ has units of rad s^{-1} T^{-1} and represents the magnetic susceptibility of the nucleus. In the case of nuclei, ω^0 is known as the Larmor frequency, expressed in either cycles per second (Hz) or angular frequency (rad s^{-1}).

The former units can be converted to the latter by multiplication by 2π. The Larmor frequency for protons in an 11.74 T field is -500 MHz.

1.2.1.1 Nuclear spin relaxation

Molecules undergo rotation, diffusion and collision with neighbours in a manner dependent on available thermal energy. Individual nuclear spin vectors are largely unaffected by such behaviour and continue to precess around the magnetic field experienced. However, since molecules are an assembly of magnetic particles (electrons and protons) whose individual magnetic fields fluctuate rapidly, the field strength and direction around which a particular nuclear vector precesses depends on the sum of the static external field and the local field, which oscillates in time and directionality. Thus the angle of precession with the external field varies over time, in a 'wandering' motion.

Crucially, this wandering is not completely random; orientations with low energy (parallel to the field) are preferred over those with high energy. A stable, anisotropic distribution of vectors is formed, which results in a small net magnetic moment (M_z) parallel to the external field ($+z$-axis) axis. The difference in population between vectors aligned parallel and antiparallel to the field is only very slight (1 in 10^4). The exponential process whereby the system 'relaxes' to this dynamic, asymmetric equilibrium is described by time constant T_1 (or rate constant, R_1).

Since M_z vanishingly small, it is difficult to measure. NMR spectroscopy observes instead magnetisation perpendicular to the field, which has to be induced by radiofrequency pulses of appropriate frequency and length[†]. When the precession angle of each spin vector is rotated by 90 ° about the x-axis, the net magnetic moment becomes aligned along the $-y$-axis, perpendicular to the external field. The spins continue to precess, and thus the net magnetic moment follows, rotating in the xy-plane at the Larmor frequency. Detected magnetisation along the $+y$-axis decays exponentially over time, with constant T_2 (or rate, R_2), and thus at time t after the pulse:

$$M_y = -M_{eq}\, cos(\omega^0 t)\, e^{-t/T_2} \tag{1.4}$$

The basis of this relaxation process is fluctuation in microscopic field strength and direction around individual spin vectors that causes the population of spins to display an increasing range of precession frequencies until there is no net magnetisation in the xy-plane (Figure 1.10a). As the spin vectors lose phase coherence, M_y decreases in amplitude (Figure 1.10b) at rate constant R_2.

[†]Radiofrequency pulses are effectively oscillating magnetic fields

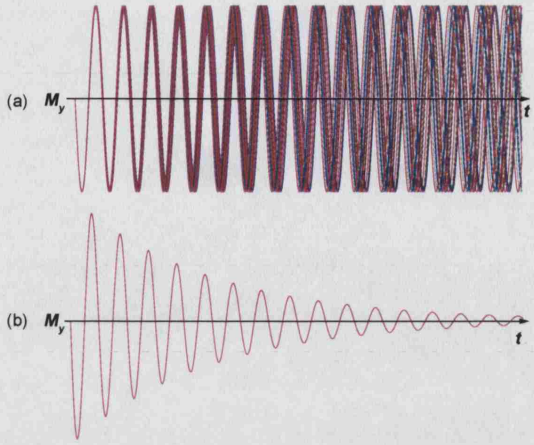


Figure 1.10 Loss of phase coherence. (a) In-phase spin vectors tend to become increasingly out of phase with one another over time, leading to an exponential decay (b) in observed signal intensity.

1.2.1.2 NMR signals

NMR spectrometers measure magnetisation perpendicular to the applied field by allowing the rotating magnetic moment to induce current to flow in a surrounding coil. After irradiation with a pulse, detected current oscillates and decays over time, resulting in a *free induction decay* (FID). In order for computers to process this oscillating current, the signal needs to be digitally sampled. Current analogue-to-digital converters (ADCs) are not able to handle frequencies the MHz range. Therefore, a receiver processes the observed signal frequency by subtracting the wave originally supplied by the radiofrequency generator (ω_{ref}). The new signal (Ω^0) fluctuates at the relative Larmor frequency:

$$\Omega^0 = \omega^0 - \omega_{ref} \tag{1.5}$$

The resultant signal, in the audio frequency range, is digitised and Fourier transformed to give the absorption Lorentzian lineshape typical of most NMR spectra. The horizontal axis displays the angular frequency (ω) of the FID waveform, with the centre of the peak positioned at the Larmor frequency, ω^0.

1.2.1.3 Chemical shifts

Recall that the Larmor frequency of a spin depends upon the product of γ and B^0. The latter component consists of a relatively static and homogenous applied field and a microscopic field. Larmor frequencies are affected by local electronic environment as well as other magnetic nuclear spins. For example, two nuclei of the same isotope within the same molecule can experience slightly different magnetic

fields and therefore precess at distinct Larmor frequencies. The resultant spectrum contains two peaks, centred at the relevant angular frequencies, shifted relative to one another. The difference in Larmor frequencies is known as the *chemical shift*, and is related to field strength in a squared fashion. Practically, it is more convenient to express angular frequencies in a field-independent manner, thus chemical shift is more commonly defined as:

$$\delta = \left(\frac{\omega^0 - \omega^0_{ref}}{\omega^0_{ref}} \right) \tag{1.6}$$

where ω^0_{ref} is the Larmor frequency of the appropriate isotope in an accepted reference compound, such as tetramethylsilane (TMS). The ratio δ remains constant over different field strengths since ω^0 and ω^0_{ref} change proportionally.

Chemical shift values are small, and are thus expressed as *parts per million* (ppm), effectively as 10^{-6}. However, δ is a highly versatile parameter, used to address a wide range of structural and functional questions. Theoretical contributions to chemical shifts are becoming increasingly well-characterised, enabling reasonably accurate interconversion between shifts and structures. For instance, the program TALOS analyses protein chemical shifts, and suggests appropriate dihedral angles[70]. Conversely, ShiftX uses a previously determined 3-dimensional structure to suggest protein chemical shifts[71].

1.2.1.4 Nuclear Overhauser effects (NOEs)

A nuclear Overhauser[72] effect describes the change in intensity of one resonance that is associated with the perturbation of equilibrium spin transitions in another. Mathematically, this is defined by Equation 1.7 in which the magnitude of the NOE, $\eta_I\{S\}$, is calculated as the relative difference between the equilibrium intensity, I_0, and that under the influence of the NOE, I. In this scheme the NOE is observed from perturbed spin S to spin I.

$$\eta_I\{S\} = \frac{I - I_0}{I_0} \times 100(\%) \tag{1.7}$$

Intensity changes induced by the NOE can be positive or negative according to the gyromagnetic ratios of the spins, the motional properties of the molecules (and therefore solution conditions) and also the spectrometer field strength. Small organic molecules exhibit small, positive NOEs whilst larger molecules display strong, negative NOEs. At intermediate molecular sizes (1000 to 2000 Da) NOEs can become very weak or disappear completely. NOE enhancements are strongly dependent on inter-nuclear distance, in an inverse r^{-6} manner. For this reason they have

become extremely useful in molecular structure determination projects, both in organic chemistry and structural biology. The perturbation of S spin equilibrium causes perturbation of I spins, which in turn may directly affect another neighbouring spin, and so on. This process, termed *spin diffusion*, results in the rapid spread of magnetisation throughout the molecule. However, this causes problems for interpretation of the NOE between S and I, which is observed to be smaller than in the case of isolated S–I spins and therefore leads to over-estimation of the inter-nuclear distance.

It is useful to distinguish between two distinct classes of NOEs – *steady-state*, and *transient*. Steady-state NOEs are measured after continuous saturation of S spins, which induces a new 'steady-state' equilibrium position in I spins. Due to the r^{-6} distance dependence, long range NOEs build up relatively slowly and require long saturation periods before becoming observable. Steady-state NOEs cannot be easily converted into inter-nuclear distances as they are influenced by neighbouring spins. In other words, the mutual steady-state NOEs between spins X and Y ($\eta_X\{Y\}$ and $\eta_Y\{X\}$) are rarely equal since the neighbours of X are not likely to match those around Y. Thus steady-state NOEs can only provide information of the relative inter-nuclear separations. However, the rate at which the NOE increases towards the new equilibrium position can be directly related to distance. This can be measured by inverting the population differences between spin states for a given resonance and observing the development of the NOE. Under these conditions, the magnitude of the NOE builds towards a maximum before receding due to spin relaxation. The NOE is therefore referred to as 'transient'. This approach is used widely in the form of NOESY experiments that provide a qualitative and quantitative description of enhancements at a particular stage in NOE development timescale. Such information can be easily converted to inter-nuclear distances for structure determination projects.

1.2.2 Ligand-based NMR screening experiments

Most ligand-based screening experiments employ ^{1}H-NMR spectroscopy due to the relatively high sensitivity and the ubiquity of protons within organic compounds. Selective observation of all compound resonances requires that spectral overlap is minimal; this can be achieved through careful library selection. Interference from receptor resonances can also complicate data interpretation although there are several solutions to this. In some ligand-based experimental schemes, spectra are recorded in the absence and presence of receptor, with ligand binding only apparent through subtraction methods. This requires highly accurate sample preparation to ensure that spectral differences are an unambiguous reflection of protein-ligand interaction

rather than changes in ligand concentration. Due to the adverse exchange rates associated with high-affinity ligands (K_d <0.1 μM), standard ligand-based screening experiments fail to generate reliable data. This disadvantage is discussed further in the context of competition-style enhancements (see Section 1.2.3).

1.2.2.1 Relaxation-edited screening experiments

Changes in NMR signal relaxation rates have been used as a probe of molecular interactions for many years. The R_2 relaxation rate is strongly influenced by the total correlation time constant, τ_c, and hence the molecular weight of a molecule. Small organic molecules display relatively slow relaxation rates whilst proteins are associated with large R_2 values. During a protein–ligand interaction, the effective τ_c of the ligand is equivalent to that of the receptor, which leads to faster relaxation of the NMR signal and reduction in the intensity of ligand resonances in the spectrum. Thus the relaxation-edited screening experiment seeks to detect changes in ligand resonance intensities recorded in the absence and presence of protein. This approach to NMR screening is discussed in more detail in Chapter 6.

In the free state, each ligand proton displays a unique R_2 value, dependent on τ_c, and other neighbouring sources of relaxation. The observed changes in R_2 on binding to a receptor are also modulated by the magnitude of chemical shift change between the free and bound states, and the rate of chemical exchange. Thus, ligand protons do not undergo equivalent changes in R_2 on binding. For this reason, it is difficult to quantify binding affinities directly from this screening experiment, although such a value is accessible by titrating the change in R_2.

Hajduk *et al.* demonstrated the screening potential of the Carr-Purcell-Meiboom-Gill (CPMG)[73] pulse sequence (relaxation-edited) using high affinity ligands for FK506 binding protein (FKBP)[74]. In this scheme, CPMG spectra were recorded in the absence and presence of FKBP, and spectral subtraction highlighted resonances that experienced changes in R_2. The identity of the ligand can be obtained directly without the need for mixture deconvolution. Protein resonances were mostly obliterated within the R_2-filtered scheme, however residual signals were removed through subtraction of a CPMG spectrum of FKBP alone. Since R_2 is dependent on τ_c, relaxation of ligands would be expected to be more efficient when screening for interaction with larger proteins. This translates into decreased protein requirements and/or shorter experimental timescales.

Spin-label enhancement (SLAPSTIC)

Jahnke *et al.* (*Novartis*) proposed a novel enhancement of the relaxation-edited approach in which the R_2 of bound ligands is dramatically increased by a nearby

unpaired electron[75]. The theoretical basis for this effect can be understood in Equation 1.8, where for two isolated, but proximal spins,

$$R_2 \propto {\gamma_A}^2 {\gamma_B}^2 \tag{1.8}$$

γ is the gyromagnetic ratio of each individual spin; for protons this is relatively small. For an unpaired electron, γ is over 650 times larger than that of a proton, and it is therefore a highly efficient source of relaxation. These species can be conjugated to proteins in the context of a 'spin-label', such as 2,2,6,6-tetramethyl-1-piperidine-1-oxyl (TEMPO). Such modifications are common-place in the field of electron paramagnetic resonance (EPR) spectroscopy.

Jahnke *et al.* proposed that spin-labels be used in two ways. First, by covalent linkage of the label to the protein near to the binding site, the relaxation rate of bound ligands could be dramatically enhanced, leading to their disappearance from the CPMG spectrum[76]. Using FKBP as a model system, the effect of spin-labelling on ligand R_2 values was explored. For a particular proton in a known ligand, R_2 values were 0.6, 0.75 and 12 s^{-1} for the free, FKBP-bound and FKBP*-bound states, respectively, where FKBP* represents the labelled protein. Thus the spin-label yielded a 16-fold enhancement in relaxation rate over the protein alone. The increased efficiency of relaxation experienced during each binding event allowed protein concentrations to be reduced from 60 to 1 μM.

The second approach suggested by Jahnke involves spin-labelling a ligand that binds to one site, and screening a library of compounds directed at a proximal site[75]. Thus compounds that bind simultaneously with the labelled ligand undergo rapid relaxation, leading to disappearance of their signals from the spectrum. Identification and conjugation of the two compounds is then expected to yield a high-affinity ligand.

The major disadvantage of both spin-labelling methods is that addition of the label to the ligand or protein could have detrimental effects on binding affinity or specificity. Fortunately, spin-labels can cause relaxation in nuclear spins up to 20 Å away, in a distance-dependent manner. Labelling strategies could therefore modify the protein without alteration of the binding site. The CPMG experiment exploits differences in observed R_2 between ligands and non-binding compounds. For small proteins (MW $<$ 10 kDa) whose relaxation rate is relatively slow, it can prove difficult to distinguish between these two species. Application of a spin-label artificially increases the difference between the free and bound relaxation rates, and enables clear identification of binders. The SLAPSTIC method is, therefore, particularly suited to screening with small therapeutic targets.

Target immobilised NMR screening (TINS)

The relaxation-edited approach to ligand screening has also been applied using immobilised receptors (TINS[77]). The method is identical to that described above for FKBP, except that the target is bound to a sepharose resin. This causes the linewidth of all test compounds to increase dramatically, however when subtracting spectra recorded in the absence or presence of target, the resonances of the ligand can still be identified. The major advantage of this technique is the ability to re-use the receptor sample many times during screening. The immobilisation of receptors also allows test compounds to be passed through the NMR sample using a 'flow-probe', which decreases the need for reloading and reshimming of the spectrometer. In principle, this method could also be used to screen against membrane-bound targets.

1.2.2.2 Diffusion-edited screening experiments

The translational diffusion coefficient of a molecule is dependent on its size and shape, and modulated by the available thermal energy and solvent viscosity. In the free state, small molecules diffuse rapidly through solution however interaction with another molecule leads to a reduction in diffusion rate, which can be detected using NMR[78]. Pulse field gradient (PFG) probe technology allows the application of a gradient of electromagnetic frequencies along the vertical axis of the sample. The Larmor frequency of spins can therefore be rendered position-dependent and used for measurement of diffusion rates.

The diffusion-edited experiment utilises a spin-echo pulse sequence in which two 90° gradient pulses flank a 180° pulse; the first pulse dephases the magnetisation along the z-axis, and the final pulse rephases the information[79]. If a molecule moves along the z-axis during the delay (Δ) between gradient pulses, then its magnetisation does not refocus completely, and its NMR signal is attenuated in proportion to its diffusion coefficient. The observed intensity I of a given resonance is therefore a function of the observed diffusion coefficient, D_{obs}, the Δ delay, and a field gradient term, K (Equations 1.9-1.11). The effective gradient K is dependent on the nuclear gyromagnetic ratio (γ), the duration of the pulse (δ) and the applied gradient strength (g). This K parameter is optimised experimentally such that signals from rapidly-diffusing small molecules are eliminated, whilst signals from protein and ligands remain in the spectrum.

$$I = I_{max}\, e^{(-K^2 D_{obs}[\Delta - \delta/3])} \tag{1.9}$$

$$\text{where} \quad K = \gamma\, g\, \delta \tag{1.10}$$

$$\text{and} \quad D_{obs} = \frac{[\text{EL}]}{[\text{L}_{\text{tot}}]} D_{bound} + (1 - \frac{[\text{EL}]}{[\text{L}_{\text{tot}}]}) D_{free} \tag{1.11}$$

There are two aspects of the diffusion-edited approach that reduce its usefulness in fragment screening. Firstly, in comparison to other ligand-based NMR screening approaches, the diffusion-edited experiment requires large amounts of protein: In order to detect changes in diffusion coefficient of a ligand on binding to a protein, a large proportion of total ligand must be associated with the receptor at any one time. To illustrate this, consider a ligand and receptor with diffusion coefficients of 20×10^{-7} and 2×10^{-7} cm^2s^{-1}, respectively, and a K_d value of 1 μM describing the interaction. When screening at equimolar protein/ligand concentrations of 0.1 mM, approximately 90 % of binding sites are occupied and the observed diffusion coefficient of the ligand is significantly reduced to 3.8×10^{-7} cm^2s^{-1} (using Equation 1.11). For weaker-affinity interactions, where K_d is 10, 100, and 1000 μM, D_{obs} is 6.9, 13.1, and 18.5 $\times 10^{-7}$ cm^2s^{-1}, respectively. Figure 1.11 indicates the effect of K_d on D_{obs} for several ligand:protein concentration ratios. For a given ligand affinity, the change in D on binding can be increased by raising the protein concentration. A 1:1 ratio of protein:ligand is found empirically to generate reliable results[74].

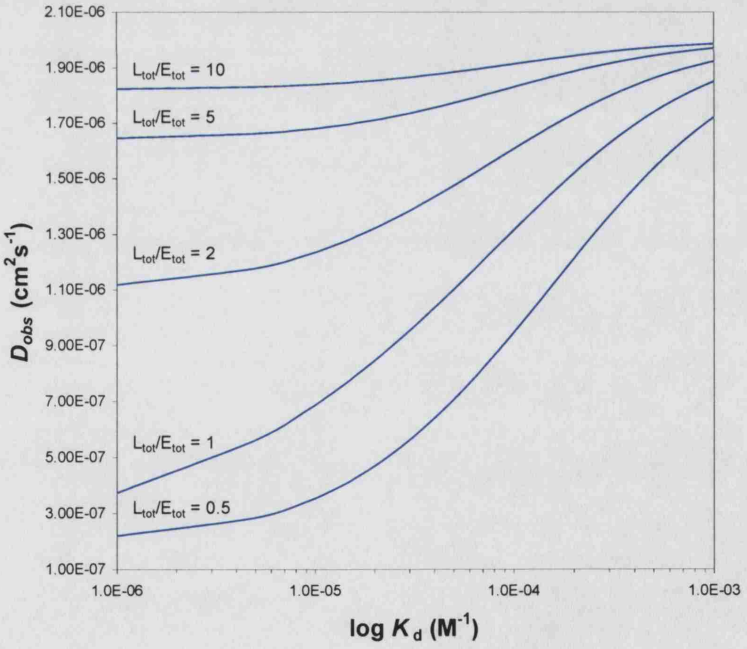


Figure 1.11 Effect of affinity on D_{obs} at ligand:protein concentration ratios of 10, 5, 2, 1 and 0.5. In this simulation, the diffusion coefficients for the protein and free ligand were 2×10^{-7} and 20×10^{-7} cm^2s^{-1}, respectively, and the ligand concentration was constant at 100 μM.

A second complication with this diffusion-edited approach relates to screening

against high molecular weight receptors. Hajduk *et al.* asserted that the sensitivity of the technique would improve with larger proteins as the difference between free and bound diffusion coefficients would increase[74]. Whilst the second half of the statement is true, there is an upper limit on receptor size appropriate to this experiment, as discussed by Meyer & Peters in their recent review[68]. At the low protein/ligand ratios employed, a large proportion of ligand molecules are bound at any one time. However, severe line-broadening associated with large proteins (> 100 kDa) causes a significant reduction in signal from ligands, leading to false-negative results in the diffusion-edited scheme. Diffusion-edited NMR screening is clearly a more robust technique for detecting binding to small or intermediate-sized receptors.

The diffusion-edited approach can be optimised for preferential selection of higher-affinity ligands. This is achieved by varying the protein concentration - clearly a large excess of protein would allow even weak binders to interact and remain in the spectrum on application of the PFG experiment. Conversely, a reduction in protein concentration increases competition between ligands with the result that weaker affinity compounds do not bind, and thus disappear from the spectrum, enabling identification of strong binders. Chen & Shapiro found that by altering protein concentrations, ligands appear in the spectrum in a sequential manner according to their affinity[80].

However they note two caveats: Firstly, they assume that all the ligands diffuse at the same rate when free and a rate equivalent to that of the protein when bound. Thus higher affinity binders tumble with the protein for longer time periods, giving the appearance of slower diffusion than weaker binders. Secondly, when signals from different compounds are compared for ranking, one must take their normal 1D signals into account. Consider a weak-binding ligand with a methyl group and a strong-binding ligand with a methine (CH) group. In the diffusion-edited spectrum, the methyl proton signal, arising from three degenerate protons, would display a higher intensity than the single methine proton, giving the impression that the former compound has the lower K_d of the two.

1.2.2.3 NOE-based screening experiments

The NOE can be harnessed for compound screening as its magnitude and phase are dependent on the molecular correlation time (τ_c). When interacting with a protein, a small molecule ligand undergoes a change in NOE behaviour – signals become strong and negative in phase. These are known as *transferred NOEs* (trNOEs). Thus comparison of 1- or 2-D NOESY spectra recorded in the absence and presence of target enables unambiguous identification of a ligand within a mixture (Figure 1.12). This analysis of intra-ligand NOEs is the preferred screening method of *Vertex* using

their SHAPES library of fragments.

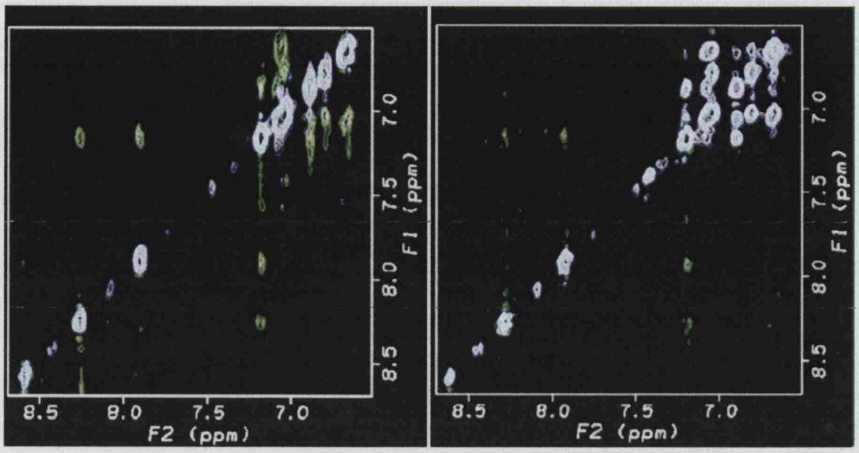


Figure 1.12 Selected region of a 2D-NOESY spectrum of nicotinic acid and 2-phenoxybenzoic acid in the absence (left) and presence (right) of p38 MAP kinase. Positive and negative contours are shown in white and green colour, respectively. A change in the phase of 2-phenoxybenzoic resonances on addition of protein indicates that this compound is a p38 ligand, whilst nicotinic acid is not. This figure is reprinted from Moore[44] with permission from *Wiley Interscience*.

Saturation Transfer Difference (STD) NMR

For many years, saturation-transfer NMR spectroscopy has been used in studies of protein–ligand complexes, however, recently it was modified for screening applications[81]. The STD method consists of two NMR experiments, recorded using the same sample, followed by subtraction to give a difference spectrum. Firstly, an 'on-resonance' pulse sequence causes selective saturation of protein ^{1}H nuclei. This is achieved by irradiating using shaped pulses at frequencies around -1 ppm, which saturates protein methyl protons. The process of spin diffusion causes magnetisation to spread throughout the protein and also to any bound ligands, within about 50-200 ms. The extent of saturation depends on the residence time of the ligand in the pocket. Release of the ligand thus transfers the saturation into solution, where it is detected.

For ligand protons involved in intermolecular NOEs with the receptor, a decrease in intensity is observed. Non-binding molecules, including buffers and impurities, show normal NOE behaviour. In order to distinguish between ligands and inert compounds, a second experiment is performed whereby the sample is irradiated at a value far from any resonance (30-40 ppm, i.e. off-resonance), yielding a normal NMR spectrum of the sample. By subtracting the on- from the off-resonance spectrum, one obtains a difference spectrum containing only signals for ligand protons. The two spectra can be recorded in an intercalated manner, with automatic subtraction,

in under 5 minutes. Protein signals can be suppressed through use of an R_2 filter (CPMG) or removed completely through use of isotopically-labelled ligands. The range of applicability of the STD technique was demonstrated by its authors[68], and the experiment has been used in several diverse situations including immobilised[82] and membrane-bound[83] proteins, and even entire virions[84].

A particularly useful aspect of the STD approach is that protons in direct contact with the protein obtain a high degree of saturation whilst distal protons can show partial saturation levels. This permits mapping of the ligand binding epitope, as demonstrated by characterisation of oligosaccharide binding to *Aleuria aurantia* agglutinin (AAA)[81] and UDP-galactose to β-1,4-galactosyl-transferase[85, 86]. Given the small size of fragment ligands, differential saturation levels may not permit unambiguous determination of binding epitopes for this class of compounds.

WaterLOGSY

Workers at *Pharmacia* developed a variant of STD NMR that utilises bound water at protein–ligand interfaces[87, 88]. Instead of irradiation at -1 ppm, the pulse sequence induces saturation at the water chemical shift ($\sim$4.75 ppm), although this also causes saturation of some H_α nuclei and protons in fast chemical exchange with water. The protein becomes saturated *via* spin diffusion and magnetisation is transferred to the ligand in the same manner as the STD method. Although some direct saturation of protein resonances occurs, WaterLOGSY is particularly suited to hydrated complexes as transfer from long-residence waters is the primary method of producing trNOE effects.

NOE Pumping

The diffusion-edited screening technique described above does not work at large ligand-protein ratios commonly employed in NMR screening[80]. Chen & Shapiro adapted the experiment such that the diffusion filter provided a means of removing small molecule resonances whilst preserving that of protein[80]. A standard NOE experiment is then performed causing magnetisation to be transferred (or 'pumped', using Shapiro's terminology) to any bound ligands, which subsequently reappear in the spectrum. As large proteins (>100 kDa) are associated with rapid spin diffusion, trNOEs should build up rapidly within the pool of ligand. However, the large R_2 relaxation rates associated with large proteins counteracts this process, inhibiting the appearance of ligand resonances in the final spectrum, possibly leading to false-negative results.

Reverse NOE Pumping

Chen & Shapiro also proposed that the transfer of magnetisation *from* bound ligands *to* the receptor could also be used as a method for screening[89]. In this scheme, receptor signals are first removed using a CPMG filter and a subsequent NOE pulse sequence permits the build-up of the weak, positive NOEs associated with small molecules. Magnetisation transferred between a ligand and the cognate receptor leads to a selective reduction in ligand resonance intensities. This can be identified by subtraction from an analogous spectrum in which signal is not lost by NOE enhancements. Interleaved acquisition of the two experiments using the same sample avoids artefacts due to temperature and magnetic field variation. Unlike the parent technique, reverse NOE pumping could be applied to screening with large proteins as the R_2 rates enhances rather than inhibits the identification of ligands.

1.2.3 Competition-based approach

A major disadvantage of all ligand-based NMR screening experiments described above is their inability to detect high-affinity ligands ($K_d < 1$ μM). The source of this problem is the slow exchange rate between free and bound ligand populations. For the experimental timescales employed (~2-4 seconds per scan), only a small proportion of slowly-exchanging ligands interact with the receptor; the remainder retain their rapid diffusion rates, slow relaxation rates and positive NOEs characteristic of freely diffusing small molecules. Differences between spectra recorded in the presence and absence of protein are thus small, causing the high-affinity ligand to be passed by – a false negative.

Groups from *Pharmacia*, *University of Georgia* and *Novartis* independently published an equivalent solution to this problem in 2002[90, 91, 92]. Dalvit *et al.* were the first to publish the application of competition equilibria to NMR screening, though admittedly papers from the other groups had been accepted for publication prior to that date (24/6/2002). They each reasoned that although it is not possible to observe direct binding of high affinity ligands, they could instead observe competition between such compounds and a reporter ligand for which binding can be robustly detected. Of the three publications, Dalvit *et al.* provided the most comprehensive introduction to the theory and application of competition kinetics to three NMR observables – linewidth, R_2 and R_1 – though in principle it could be used with any NMR observable. In the presence of protein, the reporter ligand showed broadened resonances in 1D ^{1}H spectra, indicating formation of a complex with the protein. Addition of competitor caused displacement of ligand into solution and recovery of narrow linewidths. Since the nucleus that was monitored in the reporter binds close to a phenylalanine side chain on the protein a significant difference in chemical shift

was expected for the free and bound states. Coupled with intermediate exchange rates (K_d = 7 μM), this provided a large contribution to R_2 (see Equation 1.2.2.1), and hence linewidth[‡]. Other reporter resonances did not present such a favourable situation[§].

Dalvit *et al.* subsequently used the R_2 relaxation-edited CPMG approach as described above to replicate the experiment. They also demonstrated the use of an analogous R_1-edited experiment, an approach generally not used for NMR screening as it requires selective inversion of a single resonance of each compound. However, since the competition experiment focuses on the behaviour of the reporter alone, R_1-editing can be employed.

Siriwardena *et al.* also investigated the efficacy of the competition-style approach by observing signal intensities of the reporter ligand in simple 1D proton spectra[91]. They utilised an fast-intermediate exchanging reporter (K_d = 320 μM) suitable for screening against fragments of medium-weak affinity. They also introduced a non-binding reporter molecule, which serves as an internal control for robust normalisation of data.

Jahnke and co-workers used the SLAPSTIC approach to explore competition-based screening[92]. On binding to the protein, the reporter spins are relaxed rapidly by the paramagnetic moiety, leading to their disappearance from the spectrum. Addition of a high-affinity competitor causes the reporter to remain in solution and manifest a signal. Jahnke *et al.* also noted that resonance overlap between the reporter, control and potential competitors can create difficulties in data analysis in proton spectra. They proposed use of isotope-editing to remove, for example, protons not attached to a ^{13}C nucleus (i.e. using a ^{13}C-labelled reporter). Following on from this suggestion, Dalvit *et al.* published an elegant use of fluorine nuclei (^{19}F) in which the reporter and control molecules are labelled with a CF or CF_3 moiety[93]. The natural abundance of the ^{19}F isotope is 100 %, circumventing the need for expensive isotopic labelling. A major advantage of using ^{19}F over ^{13}C or ^{15}N is its high sensitivity and large chemical shift range, which increases the effect of chemical exchange on linewidth. The technique, termed *fluorine chemical shift anisotropy and exchange for screening* (FAXS) uses the R_2 relaxation-edited experiment to distinguish between free and bound reporter molecules. A further discussion of FAXS can be found in Chapter 6. Competition enhancements of the STD-NMR[94] and WaterLOGSY[95] screening experiments have also been described recently; diffusion-based and (reverse) NOE pumping techniques have not been adapted thus far.

The major reason for application of competition kinetics to NMR screening is

[‡]The half-height linewidth of a Lorentzian peak, $\Delta\nu_{1/2}$, is equal to $1/\pi T_2$

[§]Note that as chemical shift differences are proportional to field strength, the exchange regime of a nucleus can be altered by moving to a higher field strength, with concomitant effects on $R_{2,obs}$.

to permit detection of high-affinity ligands, which would otherwise have been false-negatives. Several other advantages are also conferred, which are offset by some minor disadvantages. Competition experiments provide access to test ligand dissociation constants. This requires prior knowledge of the relationship between the fraction of ligand bound and the NMR observable. The response of the reporter ligand towards the presence of a competitor can then be related to the competition equilibria in operation, yielding a K_d value for the new ligand[69]. The affinity of the reporter can also be used to control the identification of ligands. For example, to suppress the identification of weak ligands, a relatively high-affinity reporter molecule may be employed. The competition approach also avoids detection of false-positives as only compounds that compete with the reporter cause a change in its NMR spectrum. Although this is expected to represent direct competition at one protein site, indirect allosteric mechanisms cannot be ruled out. Use of isotopically-labelled reporter compounds simplifies the NMR spectra, facilitating automatic interpretation. Finally, the reporter does not have to be drug-like – natural products, substrates, previous rejected leads can be employed.

A major disadvantage of competition screening is that ligands are not identified directly from a mixture; deconvolution experiments are required. Information regarding the binding conformation is also lost by monitoring the reporter. Interaction between the reporter and test compounds would result in false-positives.

1.2.4 Target-based NMR screening

Macromolecular diffusion coefficients, relaxation rates, and NOE characteristics are seldom modified by interaction with small molecule ligands at regularly detectable levels, unless the ligand induces considerable structural changes or protein self-association. Thus the ligand-based experimental procedures described above are not useful for screening test compounds whilst observing these NMR observables for the receptor. The only NMR observable that has been exploited for receptor-based screening is *chemical shift*: Proximity of a ligand to receptor nuclei causes a local change in electromagnetic environment, which in turn alters the chemical shift of such nuclei. If the resonance for which change is observed has been assigned to a particular nucleus in the receptor, one can locate the site of interaction. Theoretically, any NMR-active nucleus could be used to monitor chemical shift changes, however only those that are easily incorporated into a receptor have been utilised. Protons offer high sensitivity, but its relatively small chemical shift range increases resonance overlap and thus complicates data analysis. ^{13}C- and ^{15}N-labelled receptors are much more useful in this regard as certain proton resonances can be selectively observed and separated in a second spectral dimension.

The variation in chemical shift between free and bound states is entirely dependent on the electromagnetic differences between the two states. However, it is also found that the rate of chemical exchange[¶] between these states has an important effect on the NMR spectrum. For ligands with slow off-rates ($< 10\ s^{-1}$), the protein remains in either the free or bound state during the short NMR experiment, and thus separate resonances are observed, corresponding to each state. The intensity of each peak is proportional to the fractional saturation of protein sites (e.g. for the bound state resonance, intensity $\propto \frac{[EL]}{[E_{tot}]}$). Fast-exchange between states (off rate $\geq 10^4\ s^{-1}$) causes an affected protein nucleus to display a single, averaged resonance, at a chemical shift dependent on fractional saturation[‖]. For the intermediate exchange situation, a series of complex lineshapes exists, dependent on the magnitude of chemical shift difference (which is field-dependent) as well as the fractional saturation. These three exchange regimes present difficulties in the automatic analysis of 2D spectra, which is required for high-throughput, target-based screening. A further description of these three exchange regimes can be found in Chapter 4 where titration of compounds is shown to reveal differing behaviour.

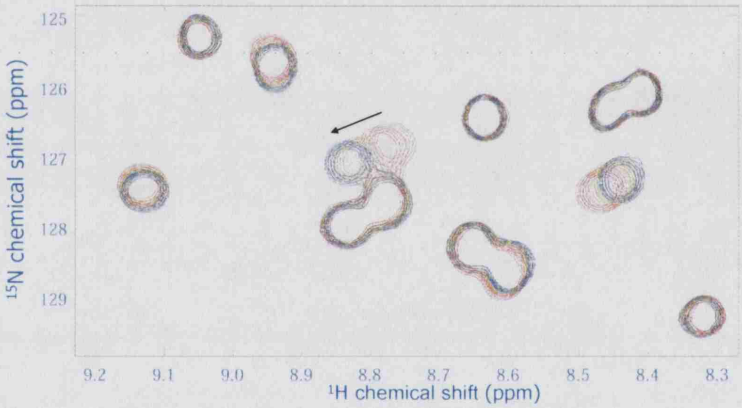


Figure 1.13 Binding by a ligand induces chemical shift changes in the NMR spectrum of the protein. In this case, phosphate causes highly specific perturbations in the backbone amide resonances of *v*-Src SH2. Only part of the HSQC spectrum is shown, for clarity.

Perhaps the most widely used 2D spectrum in protein-based NMR screening is the $^{1}H^{15}N$-HSQC (see Figure 1.13 for an example). This experiment allows selective observation of protons attached to ^{15}N nuclei, which for proteins corresponds to backbone amides and side chain amines and amides. A series of reporter nuclei are thus evenly spaced throughout the protein. The analogous ^{13}C-edited spectrum detects resonances for all ^{1}H–^{13}C groups, and is significantly more complicated. Understanding chemical shift changes in these spectra requires resolution

[¶]Chemical exchange is defined as any process in which a nucleus switches between two environments.

[‖]$\delta_{obs} = \frac{[EL]}{[E_{tot}]}\delta_{bound} + (1 - \frac{[EL]}{[E_{tot}]})\delta_{free}$

and assignment of individual resonances. For small sized proteins ($\leq$ 20 kDa) this is usually straight-forward. Medium-sized proteins (20-40 kDa) display significant resonance overlap and line broadening, which can be alleviated using TROSY[96] experiments. For proteins larger than 40 kDa, the problem of broad, degenerate resonances makes assignment very difficult to achieve. This is a serious limitation for target-based screening as many drug targets are large, multimeric complexes or membrane-associated species. Spectral complexity can be reduced to manageable levels by isotopic labelling of selected side chains (e.g. valine, leucine and isoleucine). However, this approach results in reduced coverage of the protein, compromising the ability to detect binding, especially at hydrophilic binding sites where one would not expect to find a large concentration of these hydrophobic side chains.

A major advantage of this target-based experiment for screening is that one can concentrate on the binding site of interest. Only ligands that cause chemical shift changes in specified resonances go forward as potentially useful fragments. If the atomic structure of the protein target is known, the observed chemical shift changes induced by a particular ligand can be used to locate its binding site, and predict its bound orientation. Ligands containing an aromatic moiety would be expected to cause significant chemical shift changes with magnitude and direction dependent on whether the affected nuclear spin is perpendicular to, or in the same plane as, the ring, leading to greater shielding (smaller chemical shift) or deshielding (larger chemical shift), respectively. The final advantageous feature of this target-based approach is that there is no lower limit for the dissociation constant, allowing the detection of high-affinity ligands. In fact, the experiment can be configured to avoid detection of very weak-binding fragments ($K_d > 1$ mM) by lowering compound concentration.

There are several disadvantages associated with this $^{1}H^{15}N$-HSQC screening method. Proteins that undergo large structural changes on binding a ligand will display chemical shift changes for nuclei throughout the molecule, making it difficult to find the binding site. Secondly, target-based screening requires relatively high concentrations of isotopically-labelled protein, although this can be ameliorated through investment in cryoprobe technology[97]. A necessary consequence of the elevated protein levels is high ligand concentration – typically 3-5 times that of the protein (i.e. ~0.2–1 mM), which may lead to solubility problems. Another important disadvantage of the target-based technique is the inability to identify the active ligand from within a mixture of compounds; deconvolution experiments are required.

1.2.4.1 SAR by NMR

In 1996, Fesik's group at *Abbott Laboratories* revolutionised use of NMR in the pharmaceutical industry by describing a method for effective fragment screening[39]. Although the $^{1}H^{15}N$-HSQC experiment had been used for many years in the biological sciences, Fesik *et al.* demonstrated that it could be used to generate high-affinity lead compounds. The approach, termed 'Structure-Activity Relationships (SAR) by NMR', involves the identification, through receptor chemical shift changes, of a fragment with affinity for a given protein binding site. This is optimised with regards to affinity before screening another set of fragments aimed at a second, proximal binding site, in the presence of the first fragment. Optimisation of compounds interacting with this site results in two lead fragments. At this point, the atomic structure of the ternary complex is obtained and a linker is designed on which to attach the optimised fragments, with the aim of producing a high affinity ligand. The concept of identifying, optimising, and joining fragments, using the HSQC experiment, has been patented by *Abbott*. This has likely prevented use of the approach, limiting drug design efforts. Fesik *et al.* described the identification and optimisation of fragments directed at different subsites on FKBP, which resulted in two compounds with dissociation constants of 2 and 100 μM [39]. A model of the complex was generated using NOEs between protein and ligand nuclei, which revealed the orientation and distance between the fragments. Several linkers were then designed to connect the two fragments, resulting in a range of dissociation constants, the smallest being 19 nM.

1.2.4.2 RAMPED-UP NMR

In an attempt to increase throughput, researchers at *Eli Lilly* and *Roche* pursued an approach that enabled screening of two or more protein targets at once[98]. Noting that many target resonances, being far from the binding site, do not change on addition of ligand, they sought to reduce spectral complexity by selectively labelling side chain types found in the binding site. By preparing other proteins in this manner, they were able to record HSQC spectra of a mixture of three targets. The well-dispersed resonances could be tracked on addition of known ligands, and thus they proposed use of the experiment for large scale screening of libraries against multiple protein targets simultaneously. Although in principle this technique is a valid NMR screening approach, there are several limitations to its application. The multiple proteins must not interact with one another, and additionally they must be stable and active under equivalent solution conditions. If ligands are able to interact with more than one protein then hits could be missed. For example, a

ligand in slow exchange with one protein and fast exchange with another will cause significant chemical shift changes in the former and negligible changes in the latter, unless ligand is in large excess. Thus competition between proteins can result in weak binders being missed.

1.2.5 Summary of NMR-based screening methods

The NMR screening techniques described above represent a useful, versatile and complementary approach for the detection of fragment ligands. The relative strengths and weaknesses associated with each class are summarised in Table 1.3. Selection of a particular technique will be dependent on the target in question, the expected or desired affinity of hit compounds, the information required, and relative cost. With the exception of WaterLOGSY, the ligand-based 1-D NMR experiments represent the highest throughput screening possibilities. The non-competitive 1-D approaches usually permit direct identification of a ligand from within a mixture; for competition-style and target-based experiments, deconvolution is required. Thus for primary screening of a large library (e.g. ~20,000 fragments), the ligand-based techniques may prove more efficient. Affinity data is obtainable directly from competition-style approaches, and indirectly (through titration experiments) from all others. Only NOE-based methods provide information regarding the binding epitope of the ligand, however, as discussed above, this may not be as easily obtained for fragment-sized compounds. Receptor size limitations apply in certain cases, yet other experiments, notably STD spectroscopy, can be used with immobilised, membrane-bound or other very large species. Interaction with the correct binding site can be confirmed reliably using the target-based approach, and with caution, the competitive techniques. Suspicion of allosteric displacement of a reporter ligand could be explored using protein chemical shifts. No single technique provides all the information, however in combination, the above NMR screening experiments enable rapid identification and characterisation of ligands.

1.3 Description of model system

Our exploitation of NMR in the analysis and screening of fragment ligands required selection of a suitable model system. The SH2 domain from *v*-Src was chosen for this role as it contains distinctive binding pockets capable of interaction with fragments and does not undergo conformational changes on binding. A significant body of information concerning the structural and molecular biology of SH2 domains supported this choice, including several important publications from our group, and there exists binding data for large array of natural and synthetic ligands. Addition-

NMR observable *Normal / competition style?*	Chemical shift *Normal*	NOE behaviour *Normal*				*Comp.*		Relaxation rate *Normal*		*Comp.*				Diffusion rate *Normal*
Name of experiment	Usually HSQC	STD NMR	WaterLOGSY	NOE pumping	Rev. NOE pumping	STD NMR	WaterLOGSY	T_2 Proton	T_2 Spin label	T_1 Proton	T_2 Proton	T_2 Spin label	T_2 Fluorine (FAXS)	Proton
Throughput[a]														
Time required (min)	20	10	45	5	5	10	45	10	5	10	10	5	5	10
Protein concentration (µM)	>300[b]	5	10	20	20	5	5	50	20	10	10	5	5	50
Ligand concentration (µM)	>3000[b]	100	100	500	500	100	50	50	50	100	50	10	50	50
Number of samples[c]	1	1	2	1	1	1	1	2	2	1	1	1	1	2
Number of experiments[c]	1	2	2	2	2	1	1	2	2	1	1	1	1	2
Identification of ligand ...														
... binding site?	✓	✗	✗	✗	✗	✓[d]	✓[d]	✗	✗	✓[d]	✓[d]	✓[d]	✓[d]	✗
...binding epitope?	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
... within mixture?	✗	✓	✓	✓	✓	✗	✗	✓	✓	✗	✗	✗	✗	✓
... affinity?	✓[e]	✗	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓
Range of application														
Max. receptor size limit? (kDa)	<75	✗	✗	<100	✗	✗	✗	✗	✗	✗	✗	✗	✗	<100
Receptor modification?	^{13}C or ^{15}N	✗	✗	✗	✗	✗	✗	✗	SL	✗	✗	SL	✗	✗
Lower K_D limit? (µM)	✗	0.1	0.1	0.1	0.1	✗	✗	0.1	0.1	✗	✗	✗	✗	0.1
Miscellaneous requirements	Assignments	-	H_2O[f]	-	-	RC	RC	-	-	RC	RC	RC	^{19}F RC	-
Possible enhancements	Selective labelling	P*	P*	P*	P*	R*	R*	P*	P*	R*	R*	R*	-	P*

[a] Throughput is quoted for a standard 500/600 MHz spectrometer without cryoprobe technology

[b] 3-4 fold reduction in material requirements is obtainable using cryoprobe

[c] Excluding any reference samples/spectra necessary for analysis, but recorded just once per screen

[d] Competition for protein sites by direct or allosteric mechanisms confirms site of binding in the former case

[e] Disappearance of resonances in presence of ligands indicates intermediate/slow exchange rates, therefore an indication of K_D

[f] Hydrated protein/ligand complexes are more applicable to this technique

Abbreviations: P* = Isotopic labelling of protein, R* = Isotopic labelling of reporter, RC = Reporter and control molecules, SL = Spin-labelling of protein, STD = Saturation transfer difference

Table 1.3 Comparison of NMR screening experiments.

ally, *v*-Src SH2 is readily expressed, purified and labelled (isotopically) and its small size (~12 kDa) renders it amenable to study by NMR. Therapeutic inhibitors of SH2 domain interactions have been in development for nearly 15 years. This model system is essentially identical to *c*-Src SH2, which has been validated as a therapeutic target in the treatment of osteoporosis and neoplastic disease states. Highly specific SH2 domain inhibitors are also in great need for the study of eukaryotic signal transduction.

This section provides an introduction to the structure and function of the *v*-Src SH2 domain and discusses SH2-mediated signalling specificity. The model system forms part of the non-receptor protein tyrosine kinase Src and plays important roles in signal transduction and in modulating Src kinase activation. Some human diseases are known to involve Src. The use of SH2 domain ligands to inhibit Src function is discussed, and leads into the final part of this chapter, which provides an overview of the progress in SH2 inhibitor design to date.

1.3.1 Src homology 2 domains

Src homology 2 (SH2) domains recognise phosphorylated tyrosines (pTyr or pY) embedded within specific peptide sequences on target proteins. A large number and variety of proteins are known to contain one or more copies of an SH2 domain, including non-receptor protein tyrosine kinases (PTKs, e.g. Src, Abl, Btk); protein tyrosine phosphatases (PTP-1C); lipid kinases (PI3K); lipases (PLC-γ); GTPase activating proteins (GAP); adaptor proteins (Grb2, Shc); cell structuring proteins (Tensin); and transcriptional control proteins (STATs, SPT6)*. A recent search of the SMART database† identified 1582 proteins that contain 1763 SH2 domains (~1.1 domains per protein). The taxonomic distribution of the SH2 domain is shown in Table 1.4.

Taxonomic class	SH2-containing proteins
Eukaryota	1537
Metazoa	1516
Fungi	9
Viridiplantae	3
Other	9
Viruses	45

Table 1.4 Taxonomic distribution of known SH2 domain-containing proteins (source: SMART database).

The SH2 domain is a small, independently folding module consisting of around

*PTP-1C, protein tyrosine phosphatase 1C; PI3K, 3-phosphatidylinositide kinase; PLC-γ, phospholipase C-γ; STATs, signal transducers and activators of transcription

†http://smart.embl-heidelberg.de/

100 amino acids, whose sequence and structure is conserved across the domain superfamily[99]. It consists of a central 5-stranded anti-parallel β-sheet, flanked by two parallel α-helices, with a small 3-stranded anti-parallel β-sheet immediately N-terminal to the second helix (see Figure 1.14). Overall the shape has been described as a flattened hemisphere, with the cross-sectional face forming the site of interaction with tyrosyl phosphorylated peptide sequences[100]. Both termini are located close together on the opposite face enabling the domain to be incorporated into larger proteins without affecting its ability to fold. To date, over 130 structures containing an SH2 domain have been elucidated. A large proportion of these are complexed with natural or modified phosphotyrosyl peptides.

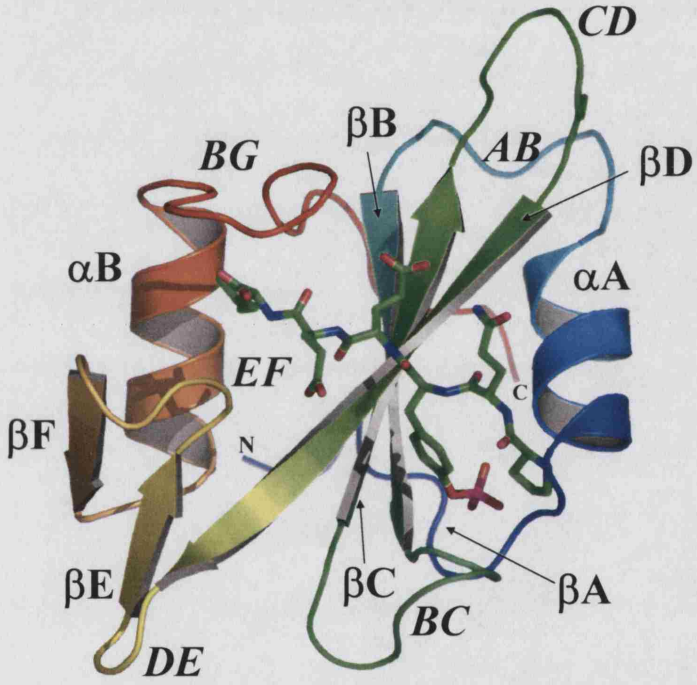


Figure 1.14 Cartoon representation of *v*-Src SH2 with bound peptide (PDB code 1SPS[101]), with structural regions labelled (loops in italics).

The peptide ligands generally bind in an extended conformation, although some SH2 domains such as Grb2 induce sharp hairpin turns in the peptide on binding[102]. Src SH2 preferentially interacts with the sequence pYEEI in a bidentate fashion (Figure 1.15). The phosphotyrosine side chain is recruited by a positively-charged hydrophilic pocket (termed the pY pocket) and isoleucine inserts into a shallower, more hydrophobic pocket (pY+3 pocket). Although the two glutamate residues are solvent exposed, they are involved in a network of hydrogen bonds with the protein, mediated by water molecules.

Phosphotyrosine itself forms a total of eight charge-assisted hydrogen bonds with side chains of R12, R32, S34, and T36 and with the backbone amide of E35 (Fig-

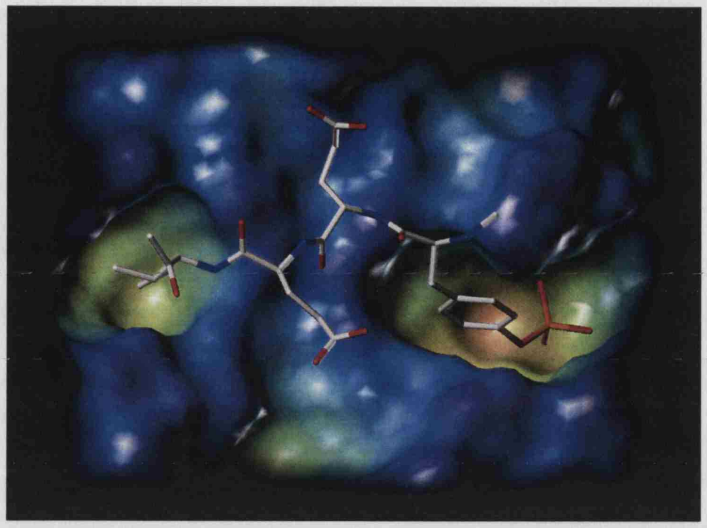


Figure 1.15 Interaction between pYEEI peptide and *v*-Src SH2. The domain surface is coloured according to cavity depth, with the pY pocket to the right and the pY+3 pocket to the left.

ure 1.16). The positively-charged ammonium and guanidinium groups of K60 and R12 also form π-interactions[‡] with the phenolic ring. A conserved hydrogen bond is observed between H58 carbonyl oxygen and the pY+1 residue backbone amide. A water molecule provides bridging hydrogen bonds between K60 H_N and pY+1 backbone carbonyl oxygen. The glutamic acid moiety of pY+1 forms a salt-bridge with K57 and van der Waals interactions with Y59. Moving further along the peptide away from phosphotyrosine, another water molecule is found to connect the guanidinium moiety of R62 with carboxylate of pY+2. The pY+3 backbone amide forms a hydrogen bond with the protein (I71 oxygen) whilst its side chain inserts into the shallow pocket making van der Waals interactions with several SH2 side chains (Y59, I71, T72, Y87, G93, L94).

SH2-peptide interactions are readily quantified by isothermal titration calorimetry (ITC), yielding thermodynamic binding data, stoichiometries, and affinities[104]. Mutational investigations of Src SH2 and peptide ligands were performed with the aim of determining individual contributions to binding affinity and specificity[105]. Phosphotyrosine itself was shown to be crucial to binding – even replacement with phosphoserine (i.e. pSEEI) results in a 10^4-fold decrease in affinity[106]. This study allocated 60 % of the ΔG_b° of pYEEI to the phosphotyrosine moiety alone[106]. For Src SH2, mutation of the most highly conserved residue (R32, Arg βB5) to alanine resulted in a ~250-fold loss in affinity. Alanine substitutions of other pY pocket residues only caused a 2-10 fold reduction; in the case of C42A (Cys βC3), a slight

[‡]These are the electrostatic interactions that occur between cationic groups and the face of aromatic rings, which display partial net negative charge (see Gallivan & Dougherty[103] for a review).

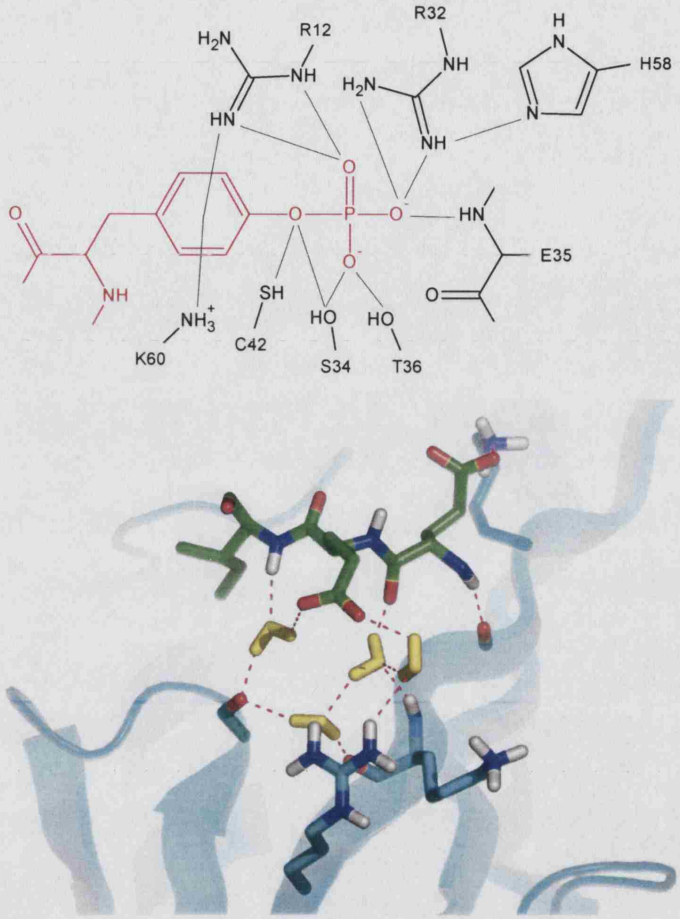


Figure 1.16 Binding of phosphopeptide by the *v*-Src SH2 domain. *Top* | Phosphotyrosine (red) makes several contacts with protein (black) side chain and backbone sites. *Bottom* | Water molecules (yellow sticks) are involved in a hydrogen bonding network (red dashes) involving protein (cyan backbone) and peptide (green backbone) atoms.

increase in affinity was observed[106]. Similar mutational investigations of Src SH2 at sites known to interact with residues C-terminal to pY indicated a maximum reduction in affinity of about 30-fold[107, 108].

1.3.1.1 SH2 domain specificity

The primary function of the SH2 domain is to recognise phosphotyrosine, as opposed to tyrosine, or any other phosphorylated compounds. For this reason, the phosphotyrosine binding site is highly conserved across the SH2 domain family, and necessarily contributes towards the bulk of the free energy of binding[2]. As a result, the free energy of binding derived from other parts of the target peptide is not large and this raises an interesting question of how signalling specificity is achieved.

Peptide sequence	K ($\times 10^6$ M^{-1})	ΔG_b° (KJ.mol^{-1})	ΔH_b° (KJ.mol^{-1})	$T\Delta S$ (KJ.mol^{-1})
EPQpYEEIPIYL	10.8	-40.1	-38.7	1.4
PQpYEEIPI	5.5	-38.5	-32.2	6.3
pYEEI	2.0	-36.8	-21.3	15.5
EPQpYQPGEN	0.07	-27.7	-25.7	2.0
pY	0.003	-19.7	-0.4	19.3
pY+1 substitutions				
PQpYDEIPI	5.3	-38.5	-37.2	1.3
PQpYQEIPI	2.1	-36.0	-31.8	4.2
PQpYAEIPI	2.9	-36.4	-32.2	4.2
PQpYGEIPI	0.16	-29.7	-20.1	9.6
pY+2 substitutions				
PQpYEDIPI	2.4	-36.4	-26.0	10.4
PQpYEQIPI	1.9	-36.0	-31.8	4.2
PQpYEAIPI	0.96	-34.3	-26.4	7.9
PQpYEGIPI	0.51	-32.6	-25.1	7.5
pY+3 substitutions				
PQpYEELPI	2.3	-36.4	-23.4	13.0
PQpYEEVPI	2.2	-36.4	-22.6	13.8
PQpYEEAPI	0.57	-32.6	-21.3	11.3
PQpYEEGPI	0.26	-31.0	-15.1	15.9
EPQpYEEVPIYL	6.3	-38.8	-28.7	10.2
EPQpYEEEPIYL	4.9	-38.1	-32.7	5.4
EPQpYEEWPIYL	3.2	-37.1	-32.2	4.9
EPQpYEEDPIYL	2.7	-36.6	-27.5	9.1

Table 1.5 Thermodynamic profile for the interaction between *v*-Src SH2 and phosphotyrosyl peptides. This table is based on that of Henriques & Ladbury[117], which draws upon data from several other publications (see references therein). All peptides were acetylated at the N-terminus and amidated at the C-terminus.

Several pivotal studies in the early 1990s analysed the affinity of SH2 domains towards a variety of natural and synthetic phosphotyrosyl peptides[109, 110, 111, 112]. High-affinity consensus sequences were identified for each domain, allowing a degree of domain classification. Some of these consensus sequences were also found in receptor tyrosine kinases known to interact with an SH2 domain[111]. A similar bioinformatic approach has been useful in predicting protein-protein interactions throughout the signal transduction proteome[113]. The sequence preference of a given SH2 domain can be perturbed by mutation of particular sites that interact with residues C-terminal to pY[107, 114, 115].

Src SH2 preferentially binds the sequence pYEEI, however the domain also interacts with a range of alternative peptides, such as pYDNV, pYTDM and pYQPG, with dissociation constants within 2 orders of magnitude of pYEEI[110, 22]. Substitution of individual residues in the pYEEI epitope caused only relatively minor changes in binding free energy, however the contributions from enthalpy and entropy were more variable (see Table 1.5 and a more extended table that I compiled for a book chapter[116]).

Despite the identification of consensus sequences, the specificity of SH2 domains towards particular peptide sequences was somewhat lower than expected. It is difficult to understand how cellular signalling *via* SH2 domains can achieve the required level of specificity with only relatively small differences in affinity between one target sequence and another. It is still generally believed that SH2-mediated signalling specificity is largely derived from residues N- or C-terminal to pY, however others have argued strongly for the additional specificity that arises through dynamic and localised formation of multimeric signalling complexes and the tandem SH2 domain arrangement observed in some proteins (e.g. ZAP70)[2, 116, 118, 119, 120]. The issue of specificity also has important implications in the design SH2 inhibitors which need to be highly selective if they are to achieve the desired change in signalling. Our current lack of understanding of eukaryotic signal transduction in general, and the precise role of SH2 domains in particular, has inhibited both the validation of these domains as pharmaceutical targets and also the development of specific inhibitors. Ironically such compounds would be extremely useful in exploring specificity further. The subject of SH2 domain specificity is discussed further in Chapter 7 in relation to the findings of this doctoral research.

1.3.2 Src-family kinases – structure, function, regulation

Since the biological role of SH2 domains is to mediate protein-protein interactions, they are almost always found in the context of a multi-domain protein. The protein SAP, which consists almost entirely of a single SH2 domain, is an interesting exception to this rule[2]. Src consists of three conserved Src Homology (SH) domains, and one unique domain (Figure 1.17). There are eight other non-receptor tyrosine kinases that are homologous in sequence to Src[§], and the group are referred to collectively as Src family kinases. These proteins are associated with the plasma membrane through myristoylation of the N-terminus.

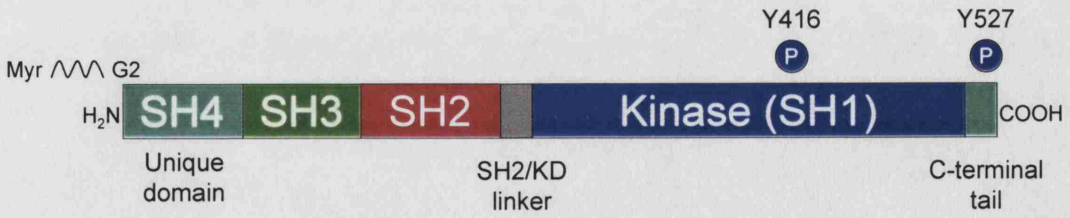


Figure 1.17 Domain arrangement within Src-family kinases. Residue numbers correspond to human Src and the figure is drawn to scale.

The enzymic activity of Src kinase is repressed in the basal state by an array of cooperative intramolecular interactions involving the SH2 and SH3 domains, and

[§]These are Blk, Fgr, Fyn, Hck, Lck, Lyn, Yes, and Yrk.

intervening linker regions[121]. Reversible phosphorylation of a conserved tyrosine side chain close to the C-terminus of the protein (Y527) plays a crucial role in regulating enzyme activity[122]. This site is a substrate for C-terminal Src kinase (Csk) and receptor-like protein-tyrosine phosphatase alpha (RPTPα), which catalyse the addition and removal of phosphate, respectively[123, 124, 125, 126].

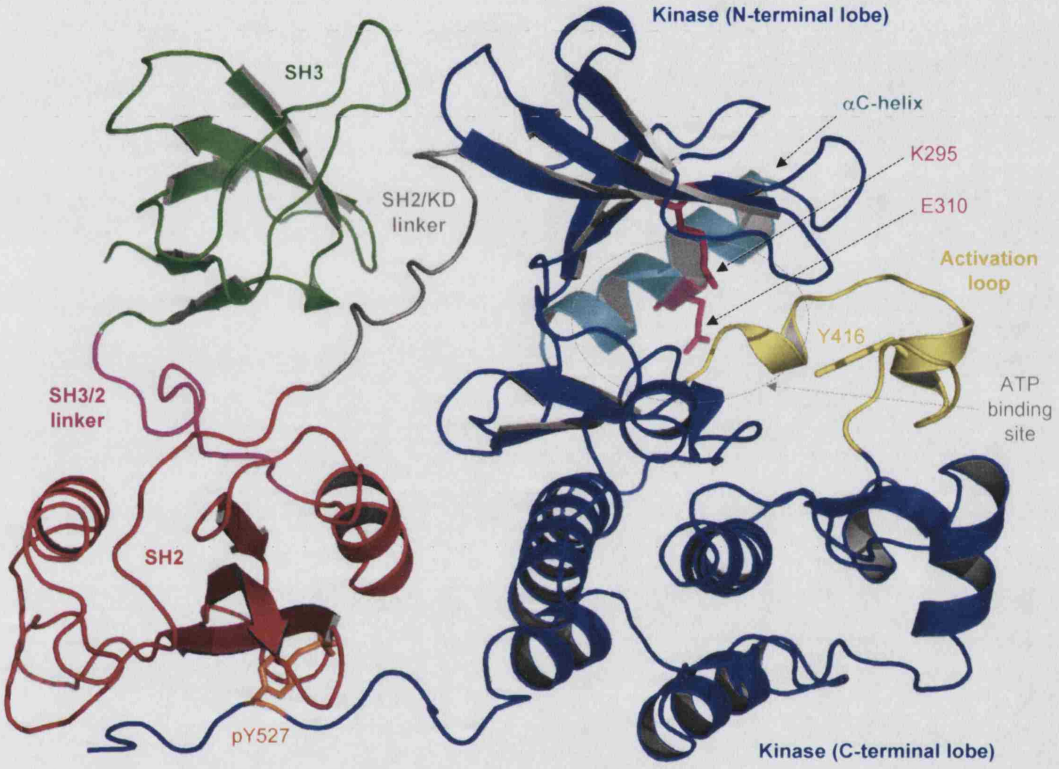


Figure 1.18 Cartoon representation of full-length Src in the inactivated state. Important residues (K295, E310, Y416 & pY527) are shown as sticks and the SH2, SH3 and kinase domains are shown in red, green and blue colour, respectively. This crystal structure corresponds to PDB entry 2SRC, with the ATP analogue omitted for clarity.

The phosphorylated sequence pY^{527}QPG interacts with the SH2 domain and restricts the C-terminal tail of the kinase domain (Figure 1.18)[127, 128, 129]. Whilst this interaction was strongly suspected prior to the elucidation of full length Src and Hck crystal structures, it was surprising to discover that the SH3 domain also makes important contacts with the region linking the SH2 and kinase domains (SH2/KD linker). The known consensus sequence for SH3 binding (**PXXP**, where X is any amino acid) is incomplete in some Src family kinases, and thus was not identified through on the basis of sequence[121]. The SH3–SH2/KD interaction induces conformational rearrangements in the kinase domain such that the substrate binding site is occluded, catalytic residues are misaligned, and the flexibility between the N- and C-terminal lobes of the kinase domain is significantly reduced.

In active protein kinases, a salt-bridge exists between a highly conserved glutamate (E310 in Src) within the αC-helix and an invariant lysine (K295)[121]. This serves to orientate the lysine for binding the phosphate groups of ATP. In the inactive state, the αC-helix is shifted, preventing formation of this crucial salt-bridge. The activation loop contains the single autophosphorylation site (Y416), which, in the inactive state, is buried between the two lobes of the kinase domain[121]. Residues at the N-terminal end of this loop also contribute to the alternative positioning of the αC-helix. A complementary interaction between a conserved leucine (L255) in the SH2/KD linker and a hydrophobic pocket in the N-terminal lobe was shown to be crucial in maintaining the inactive state – mutation of this residue to either alanine or valine leads to constitutive kinase activity, even when Y527 was phosphorylated by Csk[130]. Thus a series of cooperative and inhibitory intramolecular interactions prevent the kinase domain from becoming activated.

It is important to note that the autoinhibitory interactions of the SH2 and SH3 domains with their respective intramolecular phosphotyrosine- and proline-based ligands are sub-optimal in comparison to analogous sequences found in other signalling proteins[121]. For example, the affinity of the isolated Src SH2 domain for the C-terminal peptide pYQPG is 35-fold lower than for the activated PDGF receptor sequence, pYIpYV[22, 131]. It is not clear whether direct competition operates *in vivo*, or whether Y527 must be dephosphorylated prior to interaction with alternative species. These intermolecular binding events lead to activation of the kinase domain by a concerted disruption of the intramolecular constraints. During this process, the Y416 site in the activation loop becomes *trans*-autophosphorylated, leading to conformational changes in the N-terminal lobe of the kinase domain that confer catalytic activity[121]. This structural switch appears to alter the SH3–SH2/KD interaction such that the accessibility of the SH3 domain to alternative proline-rich sequences is increased[132].

Conversely, a decreased accessibility of the SH2 domain towards phosphotyrosyl peptides is associated with autophosphorylation. This is supported by two observations: Firstly, phosphorylation at Y416 appears to be dominant over Csk-induced phosphorylation at Y527 in both Src and Lck[133, 134]. This suggests that autophosphorylation introduces structural constraints such that pY527 cannot bind to the SH2 domain. Secondly, activation of Lyn kinase is associated with a decrease in affinity for the target peptide pYEEI[135]. Fukami *et al.* reported an interaction between a peptide sequence corresponding to residues N-terminal of the SH2 domain and the activation loop of the kinase domain[136].

To summarise, it is clear that the SH2 and SH3 domains regulate kinase activity and mediate appropriate targeting of the kinase. However, once active, the

kinase domain modulates the accessibility of the two regulatory domains; control is therefore bi-directional[130]. The atomic structure of an activated Src family kinase has not been successfully determined, probably due to unfavourable dynamics that inhibit crystallisation. A structure consisting of activated Src complexed with an upstream receptor and/or downstream target protein would be extremely useful in understanding the structural changes that occur on kinase activation and recruitment of other macromolecules. Modulators of Src activity, as well as its downstream targets, are beyond the scope of this thesis; for reviews see Brown[137] or Abram[138].

1.3.3 Src kinases and human disease

A key milestone in the history of cancer research was the discovery that one form of the disease can be transmitted without transplantation of cells. In 1911, Rous observed that injection of a filtered, cell-free extract from a cancerous chicken resulted in development of a sarcoma[¶] in healthy chickens[139]. At the time, Rous speculated that transmission of the disease was caused by a 'minute parasitic organism' or possibly a 'chemical stimulant'. Later, the causative agent was shown to be a virus, and termed the *Rous sarcoma virus* (RSV, see Martin[140] for a relevant review). This RNA-based retrovirus contains four genes termed *gag*, *pol*, *env* and *src*, corresponding to the capsid, reverse transcriptase, envelope and tyrosine kinase proteins, respectively. It is the fourth viral product, *v*-Src, that causes cellular transformation by maintaining constitutive phosphorylation of its target proteins[141, 142, 143, 144]. The discovery that *v*-Src is closely homologous to a human protein, named *c*-Src, resulted in the novel suggestion that cancer could arise by mutation of the *src* gene. This led to creation of the terms 'oncogene', whose product transforms cells (e.g. *v-src*), and 'proto-oncogene', which can express an oncogenic product when mutated (e.g. *c-src*).

Since their discovery and characterisation, genes encoding Src family tyrosine kinases have been linked to particular developmental abnormalities and cancers[145, 146]. For instance, the oncogenic potential of *v-src* is explained by the presence of a stop codon that causes premature truncation of the C-terminal region (Table 1.6). As a result, the SH2 and SH3 domains are incapable of repressing kinase activity and are themselves available for interaction with activated receptors, leading to hyper-active signaling behaviour. Similar C-terminal deletions have been detected in late stage colon cancers in humans[147]. The key autoinhibitory role of the C-terminal region of Src was underlined further on the discovery that deletion of the gene encoding Csk results in embryonic lethality in mice[148, 149].

Surprisingly, deletion of *c-src* itself leads to only one apparent phenotype – os-

[¶]A sarcoma, or sarcomata, is a tumour of connective tissue

Protein	Amino acid sequence
c-Src	FEYLQAFLEDYFTSTEPQpYQPGENL
v-Src	FEYLQAQLLPACVLEVAE--------

Table 1.6 Sequence alignment of the C-terminal region of *c*- and *v*-Src. Sequences begin at position 509 of *c*-Src and are derived from *Gallus gallus* and the Schmidt-Ruppin strain of RSV, corresponding to Swissprot[150] entries P00523 and P00524, respectively.

teopetrosis, a thickening of bone material[151]. This was unexpected as *c*-Src was not known be involved in bone physiology, however later studies showed that *c*-Src is expressed at high levels in osteoclasts, as is the related kinase Hck[152, 153]. Provision of kinase-dead *c*-Src can partially rescue the phenotype, suggesting an important role for the SH2/3 domains[154].

In normal bone physiology, there is a dynamic balance between bone resorption and deposition, which is mediated by osteoclasts. Src-deficient mice have an excess of osteoclasts, and the cells also display defects in bone resorption both *in vivo* and *in vitro*[155, 156]. The 'ruffled' osteoclast cell membrane secretes proteases, protons and other bone hydrolysing agents at the site of adhesion to bone. This border appears to be defective in the $src^{-/-}$ mice, and leads to aberrant thickening of bone material. It was thought that Src SH2 inhibitors might simulate the $src^{-/-}$ phenotype, and thus provide a treatment for bone wastage diseases such as osteoporosis, a significant cause of suffering in post-menopausal women.

Bisphosphonate compounds commonly exhibit affinity for hydroxyapatite, a component of bone material, and are used in the treatment of osteoporosis[157, 158]. *ARIAD* discovered that the 3,4-diphosphonophenyl (Dpp) fragment binds to *c*-Src SH2, thus the opportunity to target SH2 inhibitors to a specific tissue and protein presented itself. This resulted in the identification of AP22408 (1, Figure 1.19), which was active in a range of *in vitro* cellular assays and an *in vivo* model of parathyroid-induced bone resorption[159, 160, 161].

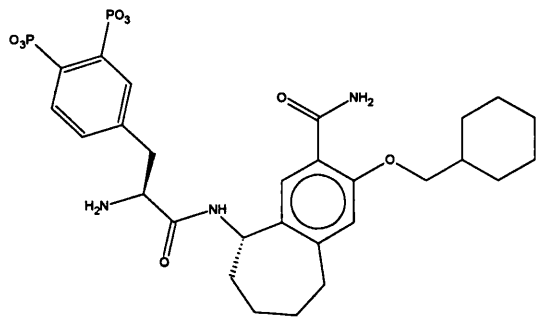


Figure 1.19 Chemical structure of AP22408, which employs the phosphotyrosine mimetic Dpp.

Deficiency of other Src family kinases results in a range of altered phenotypes. For instance, Fyn knock-out mice show neurological abnormalities[155] and Lck deletion is associated with improper Ca^{2+} signalling in Jurkat cells and hematopoietic defects in mice[162, 163]. The observation that Src family kinases are involved in both cell cycle progression and cell adhesion indicates that they might be particularly efficient proto-oncogenes[137]. In this regard, elevated levels and/or activity of *c*-Src have been detected in colon, breast and epithelial cancers (for comprehensive review see Frame[145]). Given the extensive nature of Src's downstream targets, it is not clear whether inhibitors will prove more useful in arresting cell proliferation, or tumour migration and invasion[145].

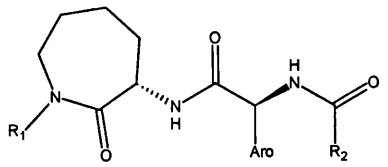
Src therefore represents an excellent example of a validated pharmaceutical target since disruption of the *src* gene is associated with a single phenotype (osteopetrosis), whilst other mutations can lead to hyperproliferative phenotypes.

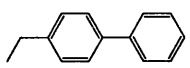
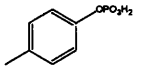
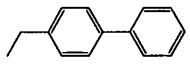
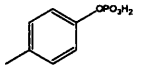
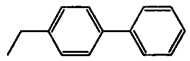
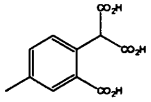
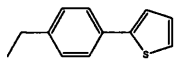
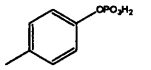
1.3.4 Inhibition of Src function

The structure and function of full length Src-family kinases provides several potential sites for targeting therapeutic inhibitors. In particular, attempts have been made to discover SH3, SH2 and tyrosine kinase inhibitors (for review see Sawyer[164]). This introduction will consider compounds targeted towards the SH2 domain only. The observation that small polypeptides and peptidomimetics are able to compete effectively with larger protein ligands for SH2 binding sites has encouraged the development of antagonists for this domain. The aim of SH2 inhibitors is to disrupt the recognition of phosphotyrosyl sequences in activated receptor proteins. For kinases such as *c*-Src, there is a fundamental problem with this approach since in the basal state, the SH2 domain interacts with the C-terminal region (pY527). Competition with this autoinhibitory site ought to lead to an increase in kinase activity; indeed this has been demonstrated *in vitro*[165, 166, 167] and in a yeast model system[168]. This is a serious consideration in the usefulness of SH2 domain inhibitors since elevated Src activity has been detected in many epithelial cancers[145].

Mandine *et al.* investigated the effect of several peptide and non-peptide *c*-Src SH2 inhibitors on kinase activity using an enzyme-linked immunosorbent assay (ELISA)[168]. Although the affinity of a given phosphopeptide for the SH2 domain was a strong determinant of the level of kinase activation, the effect of non-peptidic ligands was less clear as enzyme activity appeared to be unrelated to affinity. For instance compounds 2–5 (Figure 1.20) showed almost identical IC_{50} values, yet differential effects on Src reactivation compared to pYEEI.

This was an unexpected finding, suggestive of a complex mechanism of Src activation. The non-peptidic compounds did not directly inhibit kinase activity, nor



	R_1	R_1	Aro
2		$COCH_2CO_2H$	
3		CH_3	
4		CH_3	
5		CH_3	

Compound	SPA IC_{50} (nM)	Reactivation (%)
pYEEIE	17	100
2	12	40
3	9	100
4	4	32
5	3	25

Figure 1.20 Comparison of data for pYEEIE and synthetic ligands 2–5 that are based on the scaffold shown at the top of the figure. SH2 domain binding was measured using the scintillation proximity assay (SPA)[169]. The level of Src kinase reactivation induced by pYEEIE was normalised to 100 %.

were they less effective in binding to the SH2 domain of full length Src. Mandine *et al.* noted that Src would autophosphorylate at Y416 during the reactivation assay, and is therefore in a different state from that in the SPA binding experiment. As described above, this post-translational modification appears to lead to a decrease in accessibility for the SH2 domain and may explain why there are discrepancies between the interaction of non-peptidic ligands and the inferred 'exposed' and 'occluded' forms of the SH2 domain. However, peptide ligands caused reactivation in proportion to their affinity for SH2, and therefore appear to be insensitive to domain accessibility. This is inconsistent with the finding of Sotirellis *et al.*[135], who showed that affinity of Lyn kinase for pYEEI is dependent on phosphorylation at Y416.

Mandine *et al.* also studied the effect of the non-peptide ligands on Src reactivation within yeast (*S. pombe*) cells co-expressing Src and Csk. The results were similar to those obtained from the *in vitro* experiments – some high-affinity ligands did not lead to significant Src kinase activity. Superficially, this is not altogether unsurprising for in both cases Src was inactivated by Csk, incubated with SH2 lig-

ands and then analysed. In the yeast experiments, kinase activity was not actually measured *in vivo*, rather cells were lysed and subject to the same *in vitro* assay. Peptides were not analysed as their negative charge was assumed to preclude membrane permeation, yet this restriction is removed by cell lysis. In conclusion, Mandine *et al.* suggest that it is possible to 'block the SH2 domain without causing any Src activation' as demonstrated by 'systematic testing, both *in vitro* and *in vivo*, of SH2 binding compounds'.

There are several weaknesses in this study that prevent such bold conclusions. Firstly, the SPA binding experiments were performed using the isolated SH2 domain, and subsequently with non-phosphorylated Src, giving very similar results. This is unremarkable as in both cases the SH2 domain is accessible. Whether or not the non-phosphorylated form of Src is physiologically relevant is another matter of debate since *in vivo*, dephosphorylation at Y527 leads rapidly to autophosphorylation at Y416. Data from these experiments were then compared with results from the reactivation assay in which Src is permanently phosphorylated at Y527, and probably becomes phosphorylated at Y416 during the assay. Given the complexity of Src activation and our lack of understanding of the activated structure, this comparison may not be valid. Furthermore Mandine *et al.* offer no proof that compounds entered the yeast cell, as could easily be obtained by washing the cells repeatedly prior to lysis. No further reports of Src re-activation by SH2 ligands have been made and the issue remains unclear.

It appears almost certain that for proteins like Src, SH2 (and possibly SH3) domain inhibitors will increase the probability of phosphorylation at Y416, and result in kinase activation. However this may not present a toxicity issue as it appears that the interaction of the SH2 domain with upstream receptors is required for effective targeting of the kinase. The viral and Y527F mutant forms of *c*-Src are able to transform cells as in each case the kinase domain is constitutively active *and* the SH2 domain is accessible. Blockade of the latter domain could reverse these oncogenic effects. Further investigation into the *in vivo* effects of SH2 domain inhibitors is required to confirm this.

1.4 Progress in SH2 inhibitor design

The two principal challenges facing the design of SH2 inhibitors are replacement of phosphotyrosine and generation of specificity. Due to the high basal level of phosphatase activity within the cell, phosphoesters are unstable species and should be replaced. Moreover, peptidic amide bonds are subject to rapid protease-mediated degradation within the body and should be avoided in therapeutics. Although reduction of the negative charge of phosphate would improve cell penetration, this modi-

fication is often associated with marked losses in affinity. SH2 inhibitors have been in development for over a decade, however conversion of peptide ligands into high-affinity, specific, non-phosphotyrosyl, non-peptidic compounds has proved difficult. The affinity of inhibitors for a particular SH2 domain can be easily assessed through binding assays or biophysical techniques. The *in vivo* effects of such compounds has been much harder to establish due to specificity problems and observation of cellular response at the level signal transduction outputs (e.g. gene expression) rather than at targets immediately downstream.

ARIAD Pharmaceuticals have reported a bone-targeted Src SH2 inhibitor that shows anti-osteoporotic activity *in vivo*[159]. Potent peptidomimetic Grb2 SH2 inhibitors have been shown to block its interaction with EGF receptor, as well as arrest Ras-dependent signalling (*Novartis*)[170]. No clinical trials with SH2 domain inhibitors have been performed to date. This section will review the development of SH2 domain inhibitors, with particular focus on phosphotyrosine mimetics. Most efforts have utilised the structure-based drug design approach whereby the natural peptide ligand is adapted in an attempt to improve affinity and specificity.

1.4.1 Measuring SH2-ligand interactions and effects

The interaction between SH2 domains and cognate ligands has been quantified using various techniques. Being interested in relative potency, many industrial groups have employed biological assays in which competition between a test compound and a labelled ligand (e.g. pYEEI) is measured, yielding IC_{50} data[169, 171, 172]. Others have measured absolute affinities (NMR[173] and ITC[117]) and binding kinetics (surface plasmon resonance, SPR[172]). With a few notable exceptions, these experiments have employed isolated SH2 domains.

Scientists at *ARIAD* developed a mammalian cell 'two-hybrid' system for direct analysis of SH2 inhibitors *in vivo*[174]. In this scheme, a pY-containing sequence is fused to a DNA-binding domain whilst an SH2 domain is fused to a transcription activation domain*. Interaction between the two hybrid proteins results in expression of a reporter gene (alkaline phosphatase), the cellular activity of which is readily detected through an colourimetric assay. Addition of an SH2 domain inhibitor prevents interaction between the hybrid domains, resulting in decreased expression of the reporter gene and therefore total enzyme activity.

Due to their relatively low affinity, phosphotyrosine mimetic fragments are usually fused to a larger scaffold, such as EEI, prior to analysis. Conversely, during the optimisation of scaffolds and other modifications, phosphotyrosine is retained in order to keep affinity relatively high. Early investigations into SH2 domain inhibitors

*Or *vice versa*

relied heavily on these twin approaches, however the advent of fragment-based methods has led to independent analysis of phosphotyrosine mimetics using NMR and crystallography.

1.4.2 Phosphotyrosine mimetics

High specificity towards phosphotyrosine was demonstrated by the fact that non-phosphorylated (YEEI), and ser/thr replacement sequences (pSEEIEI and pTEEIEI) show negligible affinity[106, 175]. D-phosphotyrosine sequences were associated with only a 20–fold decrease in affinity. All high-affinity SH2 ligands published to date retain the tyrosyl ring, however many phosphate replacement motifs have been explored. A few serendipitous discoveries suggested that carboxylic acids could replace phosphate, however these are generally associated with decreased affinity compared to pY.

1.4.2.1 Phosphate replacements

As the negative charge and lability of phosphate is detrimental to the pharmacokinetic properties of phosphotyrosine-based compounds, workers have sought to replace it with other groups, as follows:

Phosphonates

Phosphonates are non-hydrolysable analogues of phosphates. Due to their higher pK_a, they are less charged at physiological pH, which improves cellular uptake[176]. Phosphonates were shown to be stable against cellular phosphatases, and 18 hours in boiling 6 M hydrochloric acid[176]. Direct replacement of phosphoester oxygen with a CH_2 group yields Pmp (4-phosphonomethylphenylalanine, 6, see Figure 1.21). Pmp-containing compounds are generally associated with significant losses in affinity compared with phosphotyrosyl analogues[177]. For example, direct comparison of phenylphosphate (7) and Pmp yielded dissociation constants of 0.3 and 5.9 mM, respectively[173]. Decreased affinity was explained as resulting from the loss of hydrogen bonding interactions with phosphoester oxygen, coupled with unfavourable increase in the pK_a of Pmp (7.1) compared with phosphotyrosine (5.7)[176].

In order to reduce the pK_a of Pmp, several fragments containing electronegative substitutions at the methylene and ring positions were designed. CH_2 protons were replaced with hydroxyl (8), F (9) or F_2 atoms (10), however only the latter compound (F_2Pmp, 4-phosphonodifluoromethylphenylalanine) produced an affinity comparable to phosphotyrosine, within a larger peptide framework[178]. F_2Pmp has been used extensively in peptidomimetics targeted at a range of SH2 domains, how-

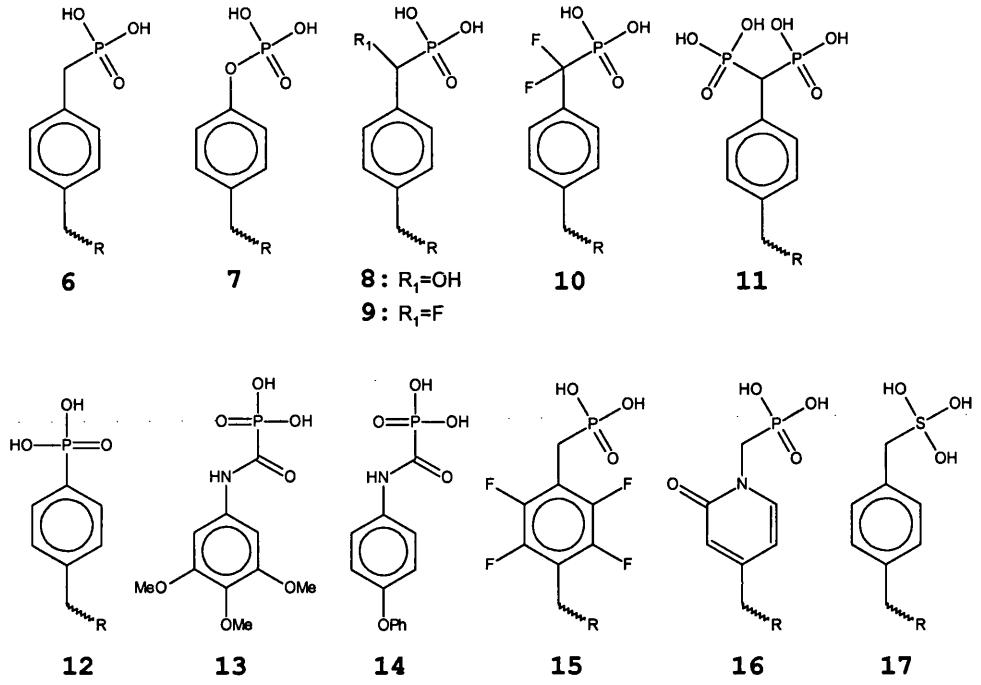


Figure 1.21 Phosphotyrosine mimetic compounds

ever such compounds were shown to be cytotoxic in whole cell assays[179]. A range of hydroxyphosphonates also showed relatively weak affinity, $IC_{50} > 10mM$[180]. Addition of a second phosphonate group to Pmp (11) resulted in an increased affinity, however only to levels observed with phenylphosphate[181]. Removal of the methylene group itself results in 4-phosphonophenylalanine (12), which also displayed moderate affinities[177]. *Aventis* validated a small range of phosphonates whereby the methylene group was replaced by an amide (i.e. Ar–N(H)–C(=O)–PO_3), and disclosed two compounds (13 and 14) with improvement in affinity over phenylphosphate[180].

Addition of four fluorines to the aromatic ring (15) resulted in a large decrease in affinity, probably because of the increased size of the ring. This tetrafluoro substitution also failed to bring about the desired decrease in phosphonate pK_a value[182]. The aromatic ring was also replaced entirely by a pyridone moiety (16), however this increased the IC_{50} value by 1000-fold, perhaps due to loss of a π-interaction with the aromatic ring coupled with a relatively high phosphonate pK_a value[183]. Addition of a second phosphonate group at the 3-position gave 3,4-diphosphonophenylalanine (Dpp, see 1 on page 68)[159]. Incorporation of Dpp into a non-peptidic analogue of pYEEI resulted in a 20-fold improvement in IC_{50}.

Sulphonates

Direct replacement of phosphorus with sulphur to give a sulphonic acid (17) resulted in an unacceptable loss of binding affinity, probably a result of an inherent increase

in size of the group compared to phosphonate and pK_a above 7[184].

Vicinal Dicarbonyls

These uncharged groups were employed to mimic the spatial positioning of certain phosphate oxygens (e.g. 18, 19, see Figure 1.22), however their use resulted in 10^2-10^3–fold increases in IC_{50} compared with phosphotyrosine[185, 184].

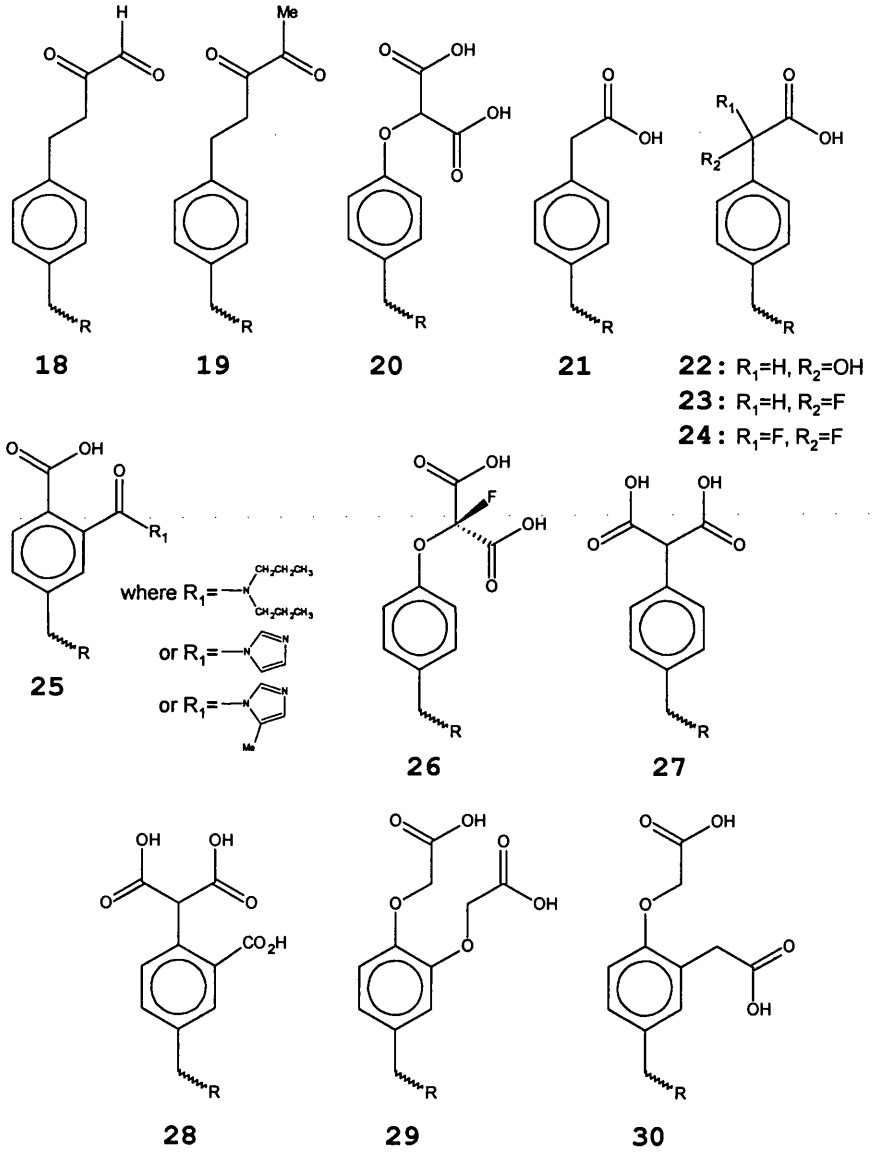


Figure 1.22 Phosphorus-deficient phosphotyrosine mimetics

Carboxylic acids

Research into another phosphate-binding protein, EPSP synthase†, revealed that malonate can act as a useful phosphate mimic[186]. Conjugation of malonic acid to

†EPSP – 5-enolpyruvylshikimate-3-phosphate synthase

tyrosine (20) resulted in an inhibitor of reduced affinity compared with phosphotyrosine. However, this reduction was in the range 8-15–fold for the SH2 domains of p85, Src and SH-PTP2 but over 100–fold for Grb2 SH2. This discovery of a specificity effect led to increased exploration of carboxylic acids as potential phosphate replacements.

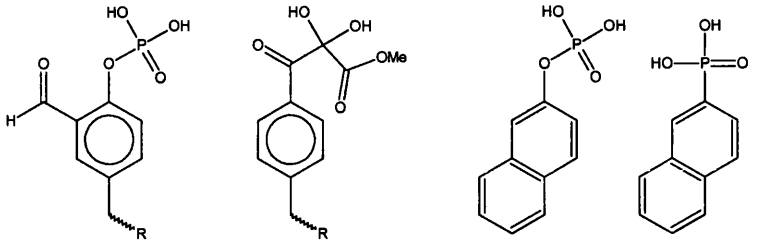
Replacement of phosphonate in Pmp with a carboxylate group yields 4-carboxymethylphenylalanine (Cmp, 21), which in fragment form showed poor affinity for Src and Lck SH2 domains, with an estimated dissociation constant greater than 10 mM (phenylphosphate was 0.3 mM)[173, 175, 184, 187]. Incorporation of Cmp into peptides also compared unfavourably with equivalent phosphotyrosyl peptides. As with Pmp, electron-withdrawing groups were added to Cmp methylene (22–24), though in contrast to Pmp, assays indicated that the F_2Cmp compound (24) had the lowest affinity, whilst the monohydroxylated form (22) was associated with the highest, but still relatively poor, affinity[184, 188, 189]. *Abbott Laboratories* screened a library of 3500 fragments against the Lck SH2 domain, highlighting the phthalamate group as a possible phosphate replacement (e.g. 25), with affinities only 3-6–fold lower than phenylphosphate[173].

The discovery that malonate and citrate interact weakly with the pY pocket resulted in attachment of multiple carboxylate groups to phenylalanine side chains. Malonyl derivatives such as *O*-2-malonyltyrosine (OMT, 20) and its monofluorinated analogue FOMT (26) displayed a 100–fold decrease in affinity compared with phosphotyrosine[190]. Removal of phenolic oxygen yielded *p*-2-malonylphenylalanine (Pmf, 27), which recovered some affinity[191]. Addition of a third carboxylate group at the 3-position (28) and incorporation into a high-affinity scaffold resulted in one of the most promising non-phosphorus mimetic to date[181]. *ARIAD* found that attachment of carboxylate groups to phenylalanine rings using ether linkages, which are stable and provide an additional hydrogen bond acceptor, resulted in compounds (e.g. 29, 30) of only moderate affinity[192].

1.4.2.2 Covalent linkage

Analysis of the complexed *v*-Src SH2 structure revealed that C42 thiol is close to the 3-position of the phosphotyrosine ring. Due to the local positive charge, the pK_a of this thiol was expected to be somewhat reduced[193]. With increased concentration of thiolate ions at physiological pH, S^- can perform nucleophilic attack upon aldehydes, forming a covalent thiohemiacetal linkage with the ligand[160, 194, 195]. This reaction is reversible and incorporation of the necessary aldehyde onto phosphotyrosine (31, Figure 1.23) resulted in the predicted protein-ligand conjugate structure[195]. *ARIAD* also detected thiol linkage with a non-phosphate dihydroxyl compound (32), which

indicated an affinity similar to pYEEI[196]. However, two chemical rearrangements lead to steady degradation of the complex over 24-72 hours.



31 32 33 34

Figure 1.23 Exploiting covalent linkage with *v*-Src SH2 (31–32) and naphthyl-based pY mimetics (33–34).

1.4.2.3 Aromatic replacements

Replacement of the aromatic moiety of phenylphosphate (IC_{50} 0.3 mM) with other mono- or bi-cyclic ring systems generally caused a decrease in affinity, except for the naphthylphosphate analogue (33), which showed a slight improvement (0.2 mM)[173, 180]. The difference in affinity between phenyl- (6) and naphthyl-phosphonate (34) ligands was greater still, with IC_{50} values of 5.9 and 1.0 mM, respectively[173]. The increased affinity of naphthyl over phenyl derivatives could be based upon extra or more optimal π-interactions with each aromatic ring.

1.4.2.4 Prodrugs

In order to facilitate cellular uptake, negative charges on phosphate and phosphonate groups can be protected using diethyl esters (see 35, Figure 1.24). These are designed to be cleaved off within the cell, enabling retention of the high-affinity phosphate moiety within the drug. Whilst improved cellular penetration was observed with these compounds, diethyl ester cleavage did not occur for the F_2Pmp analogues[177]. Mono- and bis-acyloxymethyl esters (36, 37) of F_2Pmp were more successful in this regard[197].

1.4.2.5 Miscellaneous phosphotyrosine mimetics

Attempts to replace negatively charged phosphate or carboxylate groups with a diverse range of electronegative functionalities (38-41, Figure 1.25) were not well tolerated[198]. A series of fluorinated, nitrosylated, iodinated, and hydroxylated rings showed unmeasurable affinity for Src SH2. One other nitrosylated compound (42) showed only moderate binding and also failed toxicology criteria[199].

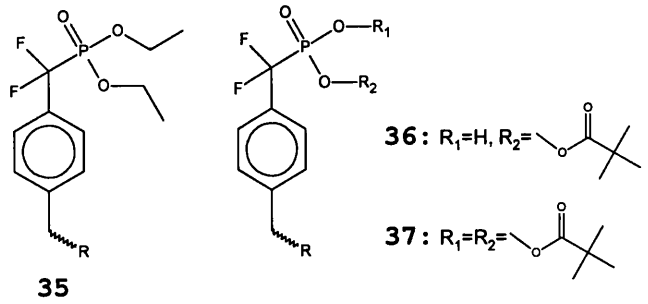


Figure 1.24 Masking of negative charge using the prodrug approach.

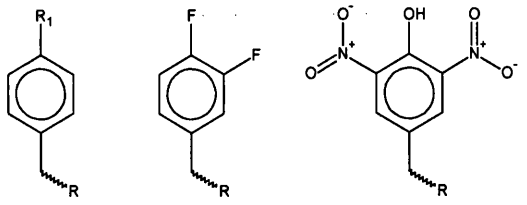


Figure 1.25 Miscellaneous phosphotyrosine mimetics

1.4.2.6 Structural analysis of pY mimetic binding

A series of x-ray crystal structures has been obtained for c-Src SH2 using a variety of fragment ligands directed towards the phosphotyrosine binding site[181]. This has enabled a comparison of the binding mode and atomic positions of each fragment. It was noted that the position of hydrogen bond acceptors across the various ligands clustered into three regions (Figure 1.26). Two such clusters are well defined and are close to the guanidinium moieties of R12 and R32. A third cluster was more disperse, interacting with backbone amides in the BC loop. The researchers proposed that a for an SH2 domain ligand to achieve high-affinity, it must contain hydrogen bond acceptors capable of making these three geometrically-conserved interactions.

1.4.3 Peptidomimetic scaffolding

An initial lack of success in replacing phosphotyrosine led to a shift in focus towards reducing the peptidic nature of the natural pYEEI ligand. Research concentrated on removing hydrolysable peptide bonds, altering side chain chemistry and addition of N- and C-terminal capping groups in order to make extra interactions and increase specificity. Some researchers explored peptide cyclisation to orient important binding groups and reduce overall flexibility, an approach that was successful in Grb2 SH2 inhibitor design.

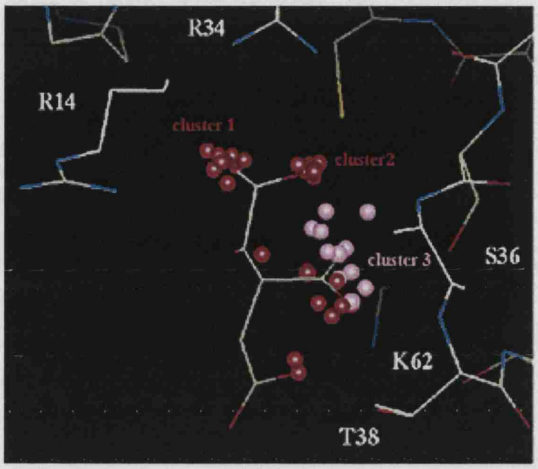


Figure 1.26 Superimposition of fragment hydrogen bond acceptors indicates three preferred locations. Citrate is shown as sticks, and is observed to match closely the required spatial positioning of acceptors. Reprinted with permission from *J. Med. Chem.* **2003**, 46, 5190. © 2003 American Chemical Society.

1.4.3.1 Non-natural amino acids

Apart from a handful of exceptions, nature predominantly uses a set of twenty amino acid side chains, and is therefore limited to the chemistry and intermolecular interactions these achieve. Modification of naturally-occurring amino acids can lead to increased affinity, and is a common first step in the peptidomimetic design process. Structural information provides the most direct route to an optimised ligand as one can model the interaction of known ligands and seek to extend the molecule to interact with novel regions on the protein surface. Lunney *et al.* introduced an amide side chain into a non-peptidic ligand in an attempt to displace a bound water molecule, and form hydrogen bond with K60 H_N and I71 carbonyl[200]. Affinity measurements indicated that the ligand (**43**, Figure 1.27) was equipotent when compared with pYEEI and a crystal structure of the complex revealed that the first two aims had been met. Unfortunately, the distance between the amide and I71 was too great to form a hydrogen bond. This amide, or close analogues, can still be found in most modern inhibitors and is a successful demonstration of rational drug design.

Another group sought to screen the effect of non-natural amino acid side chains by synthesising an array of 1000 peptides of sequence pY**X**EI and pYE**X**I where **X** is a residue with a non-natural side chain[201]. Affinity analysis of such pseudopeptides indicated optimal side chain replacements for each glutamate, which were then incorporated into the appropriate pY**XX**I motif (**44**). Yeh *et al.* also added an N-terminal capping group and achieved a 3300–fold decrease in IC_{50} value compared with pYEEI[201]. Despite this impressive feat, the three-fold problems of a negatively-charged and labile phosphate coupled with a peptidic backbone remained.

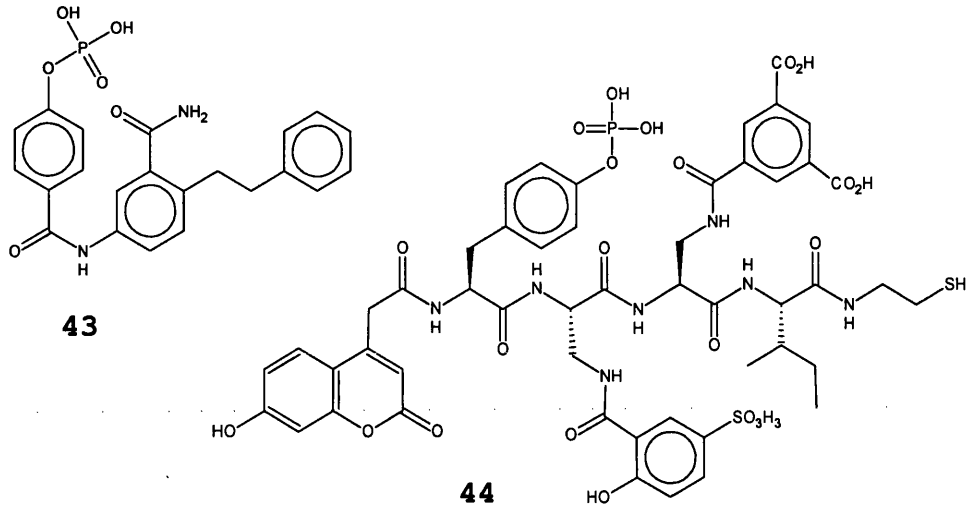


Figure 1.27 Optimisation of side chain chemistry

No structural analysis was performed using this ligand, so it is difficult to rationalise the source of binding improvements.

1.4.3.2 Capping groups

Amidation of C-termini and carboxylation of N-termini removes two charged functionalities, which can improve the pharmacokinetic properties of a peptide. Further additions can explore specific interactions with the protein surface, possibly exploiting another nearby pocket, with the aim of improving both the affinity and selectivity of a ligand. Park *et al.* fused a range of N-terminal capping groups to pYEEIE and achieved a 30–fold increase in potency for one caffeic tagged peptide (45, Figure 1.28)[202]. Lee & Lawrence noted the presence of a hydrophobic patch close to the pY pocket and subsequently screened a library of compounds containing N-terminal extensions of pYEEI[203]. They found that addition of a coumarin fragment (46) gave 2-fold improvement in K_d.

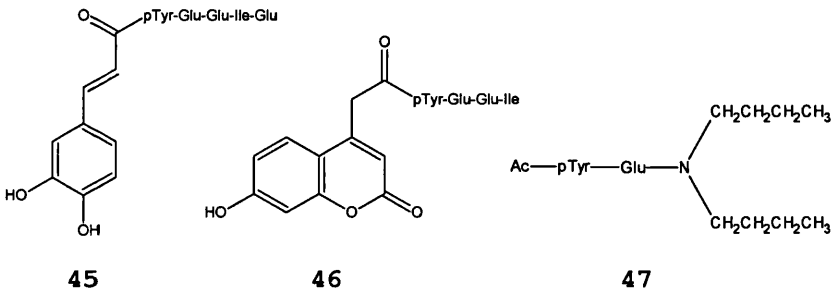


Figure 1.28 Addition of N- and C-terminal capping groups to peptide ligands

C-terminal caps were often explored in conjunction with optimising interactions with the pY+3 pocket. In one study, Charifson *et al.* joined pYE dipeptides to a variety of amino-linked aliphatic groups (e.g. 47), however these were generally associated with a 10-100–fold increase in IC_{50} compared to pYEEI[194].

1.4.3.3 Ligand pre-organisation

A ligand restrained to the bound-state configuration is expected to display a reduced entropy change on binding compared to a flexible isomer, and therefore indicate a more negative ΔG_b° value[204, 205, 206]. Whilst this appears to be a sound method of increasing compound affinity, such entropic gains are often counteracted by unfavourable changes in enthalpy, a controversial process termed entropy-enthalpy compensation[207, 208]. ITC analysis can provide overall thermodynamic profile of a system, however it has been difficult to isolate the particular source of this undesirable phenomenon. Nevertheless, other inhibitors have exploited this principle, especially enzyme transition state analogues, and there are clear examples in which nature pre-organises important side chains for functional reasons (e.g. catalytic triad of aspartate proteases, K295–E310 salt bridge in Src kinase).

Compound flexibility can be reduced by restraining mobile groups such as side chains. Ideally, this leads to configurations that are suitable for interaction with the receptor. Such modifications are not necessarily covalent – for example, hydrogen bonds were exploited in early Grb2 SH2 inhibitors in which the bound state peptide assumes a β-turn structure[209, 210]. A recent study by Nam *et al.* analysed macrocycles of pYEI tripeptide analogues and identified a 9-membered ring (48, Figure 1.29) that was 3-5 times more potent than pYEEI and a 6-membered ring (49) also produced a 5-fold decrease in IC_{50} value[211].

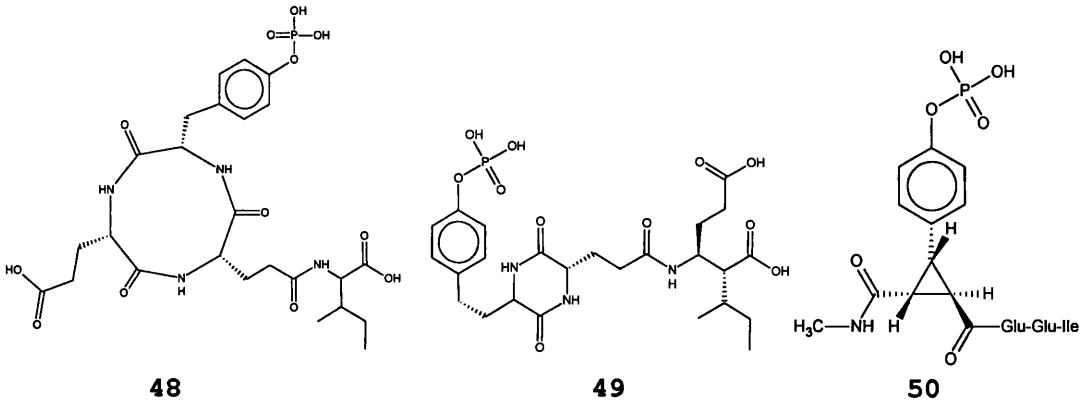


Figure 1.29 Use of ligand pre-organisation to optimise binding affinity

Davidson *et al.* used a cyclopropane group to prevent rotation of phosphotyrosine

side chain about the C_α–C_β bond (χ_1 angle)[206, 212]. Using the appropriate stereochemistry (see 50), this modification caused phosphotyrosine to adopt a bound-like conformation, confirmed in later crystal structure analysis. However, affinity measurements with both Src and Lck SH2 domains revealed no advantage over the equivalent unrestrained pYEEI ligand[212]. Calorimetric analysis showed that although unfavourable contributions to binding entropy were reduced, as desired, a compensating increase enthalpy was observed, resulting in no change in free energy, and hence affinity[206]. The source of the enthalpy difference was not located.

In 1999, *ARIAD* published results for phosphotyrosyl ligands based around a central 5/6-membered ring, which was designed to stack against the hydrophobic patch created by Y59[213, 214]. They showed that heterocycles such as thiazole and oxadiazole (51 and 52, respectively, see Figure 1.30) could be used to generate ligands of moderate affinity. *Aventis* investigated a similar range of imidazole-based compounds, although these were associated with low binding affinities[215]. Crystal structures and affinity data indicated that such cyclic groups do interact favourably with the target site, however it was concluded that a larger ring would make more efficient use of the available surface area[216]. In this regard, *ARIAD* found that a 7-membered non-planar ring (53), could match closely size and the slightly convex shape of protein surface at Y59[216]. Addition of Src-specific (thiol capture, 54) or tissue-specific (diphosphonate, 55) enhancements were subsequently reported[160, 159]. *Aventis* followed a similar 7-membered ring strategy, developing thiazapinone (56) and caprolactam (57) ligands that yielded a 15–fold increase in affinity compared with pYEEI[217, 218]. Both *ARIAD* and *Aventis* have diverted research efforts towards phosphotyrosine mimetic design, thus no further developments have been described for peptidomimetic scaffolds in the past few years.

1.4.4 pY+3 mimetics

Early studies of SH2–peptide interactions revealed that the pY+3 side chain occupies a second site, making a limited set of van der Waals interactions with hydrophobic side chains. ITC was used to measure binding thermodynamics for a range peptides with pY+3 substitutions (L, V, A, W & D)[219]. Surprisingly, alteration of pY+3 had only a relatively minor effect on affinity. Thus it appears that this secondary pocket is more tolerant towards ligand shape and chemistry than the phosphotyrosine pocket. This has been reflected in drug design efforts which have sampled a wide range of groups intended to interact with this pocket. Many research efforts have preserved the hydrophobic nature of the interaction by employing saturated/aromatic rings, or polymeric aliphatic chains, whose length are optimised for binding. For example, *Parke-Davis Pharmaceuticals* repeatedly used

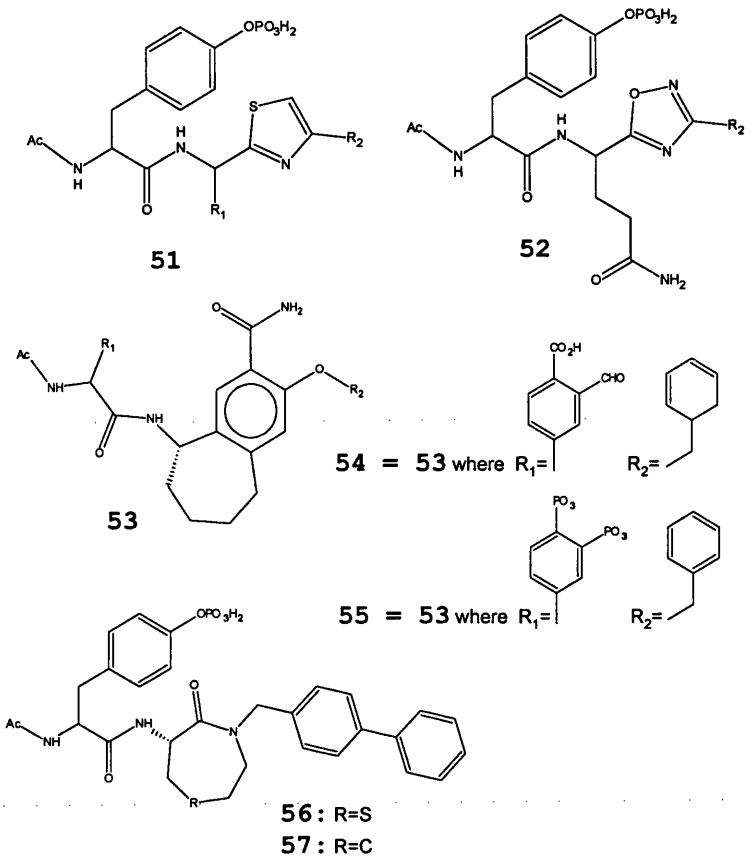


Figure 1.30 Recent non-peptide analogues that employ a core cyclic moiety.

a $(CH_2)_n$-cyclohexane moiety (58) whilst *Glaxo* designed secondary amine versions with the general formula $N(C_5H_{11})_2$ (59, see Figure 1.31)[194, 200, 220, 221].

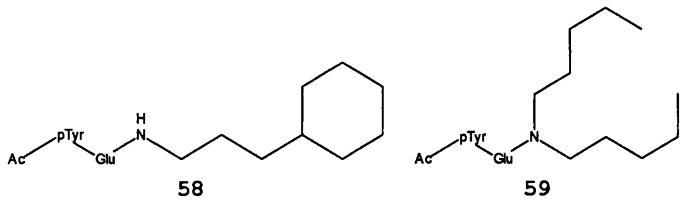


Figure 1.31 Exploring ligands for the pY+3 pocket

ARIAD used computational techniques to screen ~800 compounds directed at the pY+3 pocket. High-scoring fragments were subsequently attached to a phosphotyrosyl scaffold, known to bind with high affinity to *c*-Src SH2[214, 222]. This included range of aliphatics – linear, branched, saturated, unsaturated, cyclic – as well as many aromatic groups with minor substitutions. Finally, *Aventis* were able to show reasonable binding for bicyclic structures, in particular the naphthyl group and diphenyls[217, 218].

1.4.5 Summary of SH2 inhibitor drug design

Over the previous decade SH2 domain inhibitors have developed significantly; the most modern compounds have more favourable pharmacokinetic properties than their peptidic predecessors. *ARIAD* and *Aventis* have pioneered the development of anti-osteoporotic *c*-Src SH2 domain inhibitors, identifying useful bicyclic scaffolds from which side chains can be suitably extended. The bone-targeted compound, AP22408, displayed anti-resorptive activity while a cys-targeted analogue was selective towards *c*-Src SH2. Both inhibitors were approximately twenty times more potent than pYEEI in binding assays. Interestingly, these compounds were first described in the year 2000 suggesting that recent use of fragment-based crystallographic screening has not yet generated superior compounds.

Whilst the issue of scaffold stability has been addressed, the affinity and selectivity of the latest generation of inhibitors are not yet high enough to warrant clinical trials. The former property is being tackled by searching for high-affinity phosphotyrosine mimetics whilst the latter seeks to capitalise on differences between SH2 domain binding sites. More *in vivo* studies using existing inhibitors are needed in order to establish potency and specificity requirements, as well as confirm the effect on Src kinase activity.

1.5 Aims of thesis

The three major aims of this research were: (a) to explore the utility of NMR in the analysis and screening of fragment ligands; (b) to provide further characterisation of the model system – the *v*-Src SH2 domain – in terms of structure, dynamics, ionisation states, and the effects of ligand binding; and (c) to identify novel phosphotyrosine mimetic fragments that could be useful in SH2-related drug design.

Chapter 2

Materials and Methods

2.1 Introduction

All liquid media were prepared using distilled, deionised water (ddH_2O). Media sterilisation was achieved by either autoclaving for 20 minutes at 120 °C or use of a 22 μm filter. Unless otherwise stated, chemicals were obtained from *Sigma-Aldrich*, *BDH* and *Pharmacia*.

2.2 Preparation of labelled *v*-Src SH2 protein

2.2.1 Overexpression of *v*-Src SH2

v-Src SH2 domain protein was overexpressed in previously transformed BL21 (λDE3 plysS) *Escherichia coli* (*E. coli*) cells containing a pET3a plasmid*. Such cells were stored at -80 °C in an air-tight container. To provide individual colonies for starter cultures, cells were streaked onto Luria Broth (LB) agar plates containing 0.1 mg/ml carbenicillin and incubated for 24 hours at 37 °C. Cells were then transferred using a sterile pipette tip to 200 mL LB starter cultures containing 0.2 mg/mL carbenicillin, pre-incubated at 37 °C. The cultures were incubated with shaking (200 rpm) overnight. 50 mL of starter culture was then added to 750 mL LB containing 0.15 mg/mL carbenicillin, pre-incubated at 37 °C with shaking (200 rpm). Cell density was measured using UV/visible spectrophotometry (*CARY 100 Bio, Varian*) observing the optical density (OD) of the culture compared with an LB blank (λ = 600 nm). When the OD_{600} reached approximately 0.8 AU, overexpression of *v*-Src SH2 was induced using IPTG (final concentration = 0.35 mM). Cultures were incubated for a further 4 hours at 37 °C with shaking. Small culture samples were taken immediately prior to induction ('pre-induction') and harvesting ('post-induction') for

*This construct was kindly provided by Professor Gabriel Waksman, Birkbeck College, University of London.

future SDS-PAGE analysis. Cells were pelleted by centrifugation for 20 minutes at 5000 rpm (GS3 rotor, *Sorvall RC5B*). Pellets were then resuspended in 30 mL lysis buffer (20 mM HEPES, pH 7.5, 1 mM EDTA, 50 mM NaCl) and centrifuged for a further 30 minutes at 3000 rpm (*PK130R, ALC*) at 4 °C. The resultant pellet was resuspended in 10 mL lysis buffer per mL of cells. Protease inhibitors were then added and the samples stored at -20 °C. Pre- and post-induction samples were centrifuged for 2 minutes at 13000 rpm (*Eppendorf 5415 R*). Following aspiration of supernatant, cellular pellets were resuspended in 35 μL H_2O and 35 μL SDS loading buffer[†]. Cells were lysed by brief sonication and between 3 and 15 μL of each sample analysed by SDS-PAGE (see Section 2.2.5 for details).

2.2.2 Purification of v-Src SH2

v-Src SH2 is expressed as soluble protein, and is readily separated from bacterial proteins using affinity chromatography (*ÄKTA FPLC system, Amersham*). Frozen cells were thawed on ice and lysed using thorough sonication (5-7 times for 10 seconds at full power, on ice). Particulate matter was subsequently removed by centrifuging the lysate for 40 minutes at 18000 rpm (*SS34 rotor, Sorvall RC5B*). Clarified supernatant was then loaded at a rate of 1 ml/min onto an affinity column containing *o*-phospho-L-tyrosine–linked agarose beads (15ml, *Sigma*), pre-equilibrated in Buffer A (50 mM Tris-Cl, pH 7.5, 50 mM NaCl, 1 mM DTT). Non-specifically bound proteins were eluted by a steep gradient to 11% Buffer B (50 mM Tris-Cl, pH 7.5, 50 mM NaCl, 1 mM DTT, 1 M NaCl). A shallower gradient to 100 % B was then used to elute the specifically bound *v*-Src SH2. Buffers A and B were filtered and degassed prior to use by passing through a 22 μm filter membrane (*Millipore Express™PLUS*).

2.2.3 Concentration and dialysis

Fractions shown by SDS-PAGE to contain purified *v*-Src SH2 were pooled and concentrated in 20 mL *Vivaspin* tubes (5000 MW cut-off, *Vivascience*). Samples were then dialysed against 1-2 litres of the common buffer used in all further studies (20 mM MES, 50mM NaCl, 1mM DTT; dialysis tubing MW cut-off was 3500). Protein concentration was estimated by measurement of UV absorbance between wavelengths 240 and 340 nm, compared with that of a blank (MES buffer). The maximum signal at 280 nm was corrected by subtraction of the extrapolated baseline signal observed between 320 and 340 nm and then divided by the molar extinction

[†]SDS loading buffer contained 62.5 mM Tris pH 6.8, 2 % [w/v] SDS, 10 % [v/v] glycerol, 5 % [v/v] 2-β-mercaptoethanol and 0.001 % [v/v] bromophenol blue.

coefficient, 13730 $M^{-1}.cm^{-1}$ (obtained from ref.[223]), which, according to the Beer-Lambert law, yields the protein concentration. *v*-Src SH2 was usually stored at 4 °C at concentrations around 500 μM in the presence of 0.02 % [w/v] sodium azide to prevent bacterial and fungal growth.

2.2.4 Preparation of isotopically labelled *v*-Src SH2

Two- and three-dimensional NMR experiments require the incorporation of non-standard nuclear isotopes into the sample. To achieve uniform ^{15}N and ^{13}C labelling throughout the protein, an alternative overexpression procedure to that described in Section 2.2.1 was followed. Defrosted transformed cells were streaked onto LB-agar plates containing 0.1 mg/mL carbenicillin and 2 mg/mL glucose. Following overnight incubation at 37 °C, mature colonies were used to inoculate starter cultures consisting of 5 mL LB, 0.2 mg/mL carbenicillin, and 2 mg/mL glucose. Growth at 37 °C with shaking (220 rpm) was monitored over the next 4-5 hours and, when the OD_{600} reached 0.6 AU, 0.5 mL of the best starter culture was used to inoculate 100 mL minimal medium, which contains isotopically labelled carbon and/or nitrogen sources, as well as sufficient nutrients for bacterial growth. Instructions for preparing minimal medium can be found in Appendix A. The 100 mL culture was incubated overnight at 37 °C with shaking (220 rpm). The following morning, 10-20 mL of cell culture was used to inoculate each 500 mL minimal medium, and cells were allowed to grow for approximately six hours, when the OD_{600} reached 0.8 AU. At this point, overexpression of *v*-Src SH2 was induced by adding IPTG to a final concentration of 0.3 mM. Cells continued to grow overnight under the same conditions and, after 12-14 hours induced expression, were harvested and stored as described in Section 2.2.1. Isotopically labelled protein was purified from cell lysate in the same manner as unlabelled protein (Section 2.2.2).

2.2.5 SDS-polyacrylamide gel electrophoresis

SDS-PAGE[224] was used to assess the purity of *v*-Src SH2 samples. 12 % Tris-Tricine gel matrices were prepared (see Table 2.1) using the *Mini-Protean 3* gel electrophoresis apparatus (Biorad) with a 1 mm comb. Protein samples were boiled for five minutes, and in some cases sonicated and centrifuged, prior to loading. Electrophoresis lasted for approximately one hour, or until the solvent front began to flow out of the gel, using a constant current of 60 mA and voltages between 120 and 200 V. Proteins were then stained with a 5 % [w/v] coomassie-blue, 40 % [v/v] methanol, 7 % [v/v] acetic acid solution and destained for several hours in an aqueous solution of 20 % [v/v] methanol, 10 % [v/v] acetic acid. Biorad rainbow

markers were used as internal MW standards. Gels were preserved fully hydrated in 10 % [v/v] glycerol.

	12 % Resolving gel	4 % Stacking gel
3× Gel buffer	1.5 mL	1.0 mL
Acrylamide mix	1.85 mL	0.4 mL
ddH_2O	1.15 mL	1.6 mL
APS	100 μL	100 μL
TEMED	10 μL	10 μL

Table 2.1 SDS-PAGE recipe. The 3× Gel buffer consisted of 2.3 M Tris pH 8.75, 10.4 mM SDS, and 30 % [w/v] glycerol. The acrylamide mix was 30 % [v/v] acrylamide, 0.8 % [v/v] bisacrylamide (*ProtoGel, NationalDiagnostics*). The aqueous APS solution contained 10 % [w/v] ammonium persulphate.

2.3 Nuclear magnetic resonance spectroscopy

2.3.1 Spectrometer details

The NMR spectrometers used in this study were located at the *UCL/LICR Joint NMR Facility*, the *MRC Biomedical NMR Centre, National Institute for Medical Research (NIMR)*, Mill Hill and at *Celltech*, Granta Park, Cambridge.

UCL/LICR Joint NMR Facility

Varian UNITY*INOVA* or UNITY*plus* spectrometers, operating at proton Larmor frequencies of 600 and 500 MHz, and magnetic field strengths of 14.1 and 11.7 T, respectively, were used for the majority of protein-based NMR experiments. Both spectrometers utilise 5 mm indirect-detection triple-resonance probes with the inner coils optimised for ^{1}H detection and the outer coils calibrated to $^{13}C/^{15}N$ frequencies.

MRC Biomedical NMR Centre

The *Varian* UNITY*INOVA* 800 MHz (18.8 T) spectrometer at the *NIMR* was used to collect high resolution data for ^{15}N- and ^{13}C-edited 3D NOESY-HSQC experiments, and their associated 2D HSQC reference spectra. This work was done by Dr Mark Williams (*UCL*) and Dr Geoff Kelly (*NIMR*).

Celltech

Small molecule NMR data were collected at *Celltech* using a 400 MHz (9.4 T) *Bruker AVANCE*™. This spectrometer was used in conjunction with three different 5 mm probes: (1) $^{19}F/^{1}H$; (2) $^{1}H/^{2}H/^{13}C/^{31}P$ (Quad-nucleus probe); and (3) Broadband-tuneable (used for ^{31}P detection).

2.3.2 Acquisition of NMR data

Varian spectrometers were controlled using *VNMR* software, running on *Sun Microsystems*™) workstations, with all experiments performed at 25 °C, unless otherwise stated. *XWIN-NMR* software was used to operate the *AVANCE*™ spectrometer, running on a *Silicon Graphics Inc. (SGI)* workstation. All experiments on the *Bruker* system were done at 21 °C. All NMR samples contained approximately 10% [v/v] D_2O, which acted as a frequency lock to correct for the drift of magnetic field with time. Standard 5mm NMR tubes (*Wilmad Glass Co., Inc.*) were used for all screening experiments. Certain experiments were run on smaller samples in *Shigemi* tubes (*BMS2/BMS3, Shigemi Co., Ltd.*). Signals from water were suppressed using the presaturation method, except on the 800 MHz spectrometer where the 'Watergate' method was employed. Unless otherwise stated, all NMR experiments employed the same buffer (20 mM MES, 50 mM NaCl, 1 mM DTT, 10 % [v/v] D_2O), referred to throughout as 'NMR buffer'.

The set of pulse sequences and spectral parameters used to obtain resonance assignments and geometric restraints for the calculation of the *v*-Src SH2 solution structure are summarised in Table 2.2.

Experiment	^{1}H			D2			D3			Mix[d]
	SF	SW	NP	Nuc	SW	NP	Nuc	SW	NP	
HSQC[b]	600	8000	2048	^{15}N	2400	256				
CT-HSQC[c]	800	10500	3150	^{13}C	16000	128				
CT-HSQC[a]	500	4700	1024	^{13}C	3770	64				
CBHD[a]	600	6000	1024	^{13}C	4800	38				
HNCA[b]	600	4440	1024	^{13}C	5700	72	^{15}N	1900	36	
HNCO[b]	500	3700	1024	^{13}C	1625	64	^{15}N	1650	32	
HNCACB[b]	500	3700	1024	^{13}C	7750	70	^{15}N	1650	32	
CBCACONH[b]	600	4440	1024	^{13}C	9000	60	^{15}N	1900	36	
HCC(CO)NH[b,c]	500	3700	512	^{13}C	8500	68	^{15}N	1650	36	
HNHA[c]	600	7500	1024	^{1}H	5700	64	^{15}N	1430	32	100
HNHB[c]	600	7500	1024	^{1}H	5700	64	^{15}N	1430	32	100
NOESY-HSQC[c]	800	10500	1024	^{1}H	10500	194	^{15}N	1905	48	200
NOESY-HSQC[c]	600	4440	434	^{1}H	7500	139	^{15}N	1430	36	50
TOWNY-HSQC[c]	600	4440	512	^{1}H	7500	128	^{15}N	1430	32	46
NOESY-HSQC[c]	800	9350	912	^{1}H	7050	60	^{13}C	9350	164	125
NOESY-HSQC[a]	500	4700	1024	^{1}H	5700	72	^{13}C	3770	40	100
HCCH-TOCSY[c]	600	4500	1024	^{1}H	4500	64	^{13}C	9800	72	19
HCCH-TOCSY[c]	600	4500	1024	^{1}H	4500	64	^{13}C	9800	72	6
HCCH-TOCSY[a]	500	4700	1024	^{1}H	2250	48	^{13}C	3770	40	6

Table 2.2 NMR parameters employed for experiments. CT-HSQC, constant-time HSQC. [a]Spectra centered at aromatic proton region; [b]Triple resonance spectra; [c]Spectra used to assign side chain resonances; [d]Mixing time.

The ^{1}H^{15}N-HSQC pulse sequence was also employed in the analysis of *v*-Src SH2 hydrogen bonding (Section 3.9.3), ligand binding (Sections 4.1 & 6.6) and

side chain ionisation states (Section 4.3). Such experiments employed ^{1}H and ^{15}N sweep widths of 4000 and 1650 Hz, respectively, with the offset fixed at 1100 Hz. Unless otherwise stated, these spectra were recorded using 80 transients in proton and 40 increments in the nitrogen dimension, requiring approximately 25 minutes of acquisition time. ^{1}H^{15}N-HSQC experiments were performed using the 500 MHz spectrometer except for those used in the structure determination process where 600 and 800 MHz spectrometers were employed.

For the investigation into side chain protonation states (Section 4.3), two types of constant-time ^{1}H^{13}C-HSQC spectra were obtained; one was used to observe the aliphatic region and the other was centred on the aromatic region, corresponding to offset frequencies of -160 and -655 Hz. For the aliphatic spectrum, ^{1}H/^{13}C sweep widths were 6000/4500 Hz, collecting 64 transients in the proton dimension and 192 scans in the carbon dimension (11.5 hours acquisition time). The aromatic constant-time ^{1}H^{13}C-HSQC employed ^{1}H/^{13}C sweep widths of 4700/3770 Hz, 64 transients, and 96 scans (5 hours acquisition time).

2.3.2.1 NMR relaxation data

R_1, R_2, and HNOE relaxation experiments were used to analyse the backbone mobility of *v*-Src SH2 (see Section 4.2). FID data from R_1/R_2 experiments were recorded as an array of 2D planes, with each plane corresponding to a particular relaxation delay determined by the number of cycles (ncyc) around the pulse sequence loop (see Table 2.3). One plane in each experiment was repeated to enable error analysis. Cross-relaxation NOE data were recorded in a similar pseudo-3D manner with planes corresponding to the saturated state and the non-saturated state. Noise levels were obtained using *Plot2*[‡] and peak heights for each plane were extracted from *ANSIG* using scripts written by Drs M. Williams and M. Pfuhl.

Experiment	Basic relaxation delay (s)	ncyc parameter array
R_1	5.0301×10^{-3}	2,8,16,32,64,128,192,256,320,32
R_2	33.216×10^{-3}	1,2,3,4,5,6,7,8,9,10,3

Table 2.3 Relaxation delays used in R_1 and R_2 experiments. In each case, the basic relaxation delay (Δ) is calculated as follows: For R_1, $\Delta = 0.005 + 2 \times \text{pwhn} +$ WFG where pwhn is 12.15 μs and WFG is 5.75 μs; for R_2, $\Delta = 32(\text{pwn} + \tau_T)$, where pwn is 38 μs and τ_T is 450 μs

[‡]*AZARA*, v2.7, © 1993-2002 Wayne Boucher and Department of Biochemistry, University of Cambridge.

2.3.3 Processing and analysis of NMR data

Mathematical processing of free induction decay (FID) data from *Varian* spectrometers was performed using the *NMRPipe Spectral Processing and Analysis System* and viewed using *NMRDraw*[225]. Chemical shifts were referenced indirectly to sodium 2,2-dimethyl-2-silane-pentane-5-sulphonate (DSS). Two dimensional heteronuclear data were processed in the following typical manner: The ^{1}H dimension time-domain data were treated using an apodisation function (e.g. SP, POLY or GMB) and subsequently zero-filled, Fourier transformed and phased appropriately. The second (indirectly-detected) dimension was usually linear predicted, and then treated in the same manner as the ^{1}H dimension. Baseline correction was achieved using the POLY function. Three-dimensional FID data were processed in a similar manner, except that the ^{1}H dimension data were temporarily circular shifted in such a way that the low-frequency water signal was located in the centre of the frequency window to allow efficient solvent suppression using an appropriate apodisation function (e.g. SOL). For 3D spectra, linear prediction was applied to the third dimension rather than the second. Although the relaxation data (see Section 4.2) were recorded as pseudo-3D spectra, the individual planes were extracted and processed in an identical fashion, as described for the 2D data above. NMR data processed by *NMRPipe* were converted to *AZARA* format and analysed using *Plot2* (part of *AZARA*) and *ANSIG* v3.3[226]. Peak positions were automatically detected using the *AZARA* 'peakfind' macro, and subsequently edited by hand to alter or remove unlikely crosspeaks.

2.4 Isothermal titration calorimetry

ITC is a universal technique for determining the thermodynamic profile of chemical processes, such as ligand binding[227, 228]. Molecular recognition events involve a transfer of heat energy either to or from the surroundings (exothermic or endothermic, respectively). The titration calorimeter is able to measure these heat transfers and such data can be used to quantify the change in enthalpy, ΔH_b°, the binding constant, K ($=1/K_d$), and the stoichiometry, n. Using Equation 2.1, the overall change in free energy, ΔG_b°, and entropy, ΔS_b°, can also be derived[§].

$$\Delta G_b^\circ = -RT \ln K = \Delta H_b^\circ - T\Delta S_b^\circ \tag{2.1}$$

ITC experiments are performed at a constant temperature by titrating one molecule, usually the ligand, into a sample cell containing the other. Ideally, the

[§]R is the gas constant and has a value of 8.314 J.K^{-1}.mol^{-1}

titration should reach a state of saturation in order to obtain accurate limits to the sigmoidal curve. The small amount of heat associated with ligand injection and dilution is measured in a second experiment in which the ligand is titrated into buffer alone. These heats are subtracted from those measured for the protein/ligand titration and the corrected data is then fitted to an appropriate binding site model, for example, single-site binding.

The binding constant window that can be determined reliably by ITC is defined by the unitless c value, which is the product of P_{tot}, the total macromolecular concentration, and K. c-values in the range 1–1000 provide data that can be fitted to equations describing binding equilibria. For molecular interactions outside of this range (i.e. very low or high affinity), ITC may still be used to generate thermodynamic information through competition experiments[229, 230]. These were employed in Chapter 6 to study the binding of fragment ligands for the SH2 domain and the concepts are described there (Section 6.2.2). All ITC experiments were performed using a VP-ITC calorimeter (*Microcal*, Northampton, MA) at 25 °C in 20 mM MES pH 6.00, 50 mM NaCl, and 1 mM DTT. Thermodynamic data were analysed using *ORIGIN (OriginLabs)*.

Chapter 3

Solution structure of *v*-Src SH2

3.1 Introduction

This chapter describes the determination of the apo solution structure of *v*-Src SH2 through application of standard NMR techniques and an automated NOESY assignment and structure calculation procedure. A separate investigation into the temperature dependence of backbone amide chemical shifts is also reported, which contributes towards our understanding of hydrogen bonding within the domain.

Ongoing studies of the *v*-Src SH2 domain in several areas, such as the detection and quantification of ligand binding, molecular dynamics simulations, and NMR relaxation analysis require detailed understanding of the atomic structure of the unliganded domain in solution conditions. Two crystal structures of the *v*-Src SH2 domain were published a decade ago by Waksman *et al.*, corresponding to the peptide-free (1SPR*) and peptide-bound (1SPS) forms[101]. The former structure does not accurately represent the apo state since the domain was crystallised in 0.2 M K_2HPO_4, which interacts weakly with the phosphotyrosine pocket. Indeed, electron density corresponding to a phosphate ion was found in MolA of the asymmetric unit, while the other three molecules displayed side chain orientations indicative of bound ligand. Heavy atom replacements to determine x-ray diffraction phase made use of 100 mM iridium chloride, which was also found in the pY pocket, and 1 mM mercuric chloride, found in the vicinity of cysteines[101]. Finally, the inter-domain contacts observed for MolA/MolC and MolB/MolD pairs within the asymmetric unit of the crystal also do not correspond to solution conditions.

A second determination of the *v*-Src SH2 structure in the peptide-free state (1BKL) employed ammonium sulphate buffer, but again this anion was found in the pY binding site[232]. Thus the only available structures of *v*-Src SH2 do not properly describe the apo solution state. Other Src family SH2 domain structures

*Protein data bank (PDB)[231] accession code.

have been determined in a supposed 'apo' state, however these were also obtained in the presence of weak-affinity ligands (see Table 3.1).

SH2 domain	PDB	Method	Conditions
Blk[233]	1BLJ	NMR	10 mM phosphate
Fyn[234]	1G83	NMR	50 mM phosphate, 100 mM sulphate
Hck[235]	3HCK	NMR	50 mM phosphate
Lck[187]	1BHH	X-ray	100 mM citrate, 200 mM acetate
v-Src[232]	1BKL	X-ray	2 M ammonium sulphate
v-Src[101]	1SPR	X-ray	0.2 M phosphate

Table 3.1 Src family SH2 domain structures determined to date were buffered by anions that interact with the pY binding site.

It was expected from the outset that the overall fold of the solution and crystal structures of *v*-Src SH2 would be very similar. Both structure determination projects used essentially identical protein sequences (the crystal structure does not include C-terminal residues 105 & 106). Any differences between the solution and crystal structures must be explained by the experimental conditions employed (Table 3.2), the most important distinction being exclusion of phosphate ions from the NMR sample.

	x-ray	NMR
v-Src SH2	20 mM	0.5 mM
Buffer	10 mM HEPES	20 mM MES
pH	8.0	6.0
Salt	0 mM	50 mM NaCl
Other	5 mM DTT	1 mM DTT
	200 mM K_2HPO_4	0.02 % [w/v] NaN_3
	5 mM EDTA	
Temperature	RT	25°C

Table 3.2 Experimental conditions used for determination of peptide-free x-ray crystal (1SPR[101]) and the apo solution (NMR) structures.

The secondary structure prediction from chemical shift information (see below) was consistent with the expected homology between the crystal and solution structures. This permitted use of some information from the crystal structure during NOE assignment in order to identify and exclude data that was clearly inconsistent with the overall fold of *v*-Src SH2. Further iterations of automatic assignment and structure calculation were performed without reference to the crystal structure.

Structure calculations for the *v*-Src SH2 domain utilised inter-atomic distance restraints derived from NOEs and predicted hydrogen bonds, as well as angle restraints in the form of empirical coupling constants and predictions from chemical shift indices. The triple-resonance 3D experiments were used to assign backbone resonances (N, C′, C_α, H_N), which enabled prediction of secondary structure and

backbone dihedral angles. Heteronuclear 3D TOCSY and NOESY experiments were then used to assign side chain nuclei to a high degree of completion. Coupling constant data that describe potential phi (ϕ) backbone angle ranges were extracted from the HNHA experiment. NOESY spectra were automatically assigned using the chemical shift data and used to calculate structures in an iterative manner by the program *ARIA*[236]. Repeated applications of *ARIA* enabled refinement of the automatically picked NOE peak list. Hydrogen bonds were identified in late stage structures and specified as distance restraints in further calculations.

3.2 Sample preparation, data acquisition and processing

Isotopically-labelled *v*-Src SH2 samples were prepared as described in Section 2.2.4. Experiments were carried out using one ^{15}N-labelled and one ^{13}C/^{15}N double-labelled sample, at protein concentrations between 0.5-0.6 mM. NMR data was collected using three spectrometers, operating at 500, 600 and 800 MHz (for details see Section 2.3.1) using the experimental parameters listed in Table 2.2 on page 89. Free induction decay (FID) data sets were processed using *NMRPipe* scripts, as described in Section 2.3.3.

3.3 Resonance assignment

The sequence specific assignment of nuclear resonances utilised established methods. Backbone resonances were assigned by logical interpretation of triple-resonance spectra and this information provided a basis for assignment of side chain resonances. Published characteristic chemical shift data were used to facilitate residue-specific resonance assignments, however this was verified in each case by resonance connectivity in TOCSY spectra. Virtually complete assignments were obtained for all residues and submitted to the BioMagResBank[†] database under accession number 6503 (to be released with associated publication).

3.3.1 Backbone resonance assignment

The purpose of backbone resonance assignment is to generate a list of chemical shifts for protein backbone nuclei (N, C′, C_α, H_N). Several triple-resonance[‡] spectra were used to assign backbone nuclear resonances in the following manner:

^{1}H and ^{15}N chemical shift coordinates for a backbone amide (residue i) are used to locate the relevant position in the 3D triple resonance spectra. HNCA and

[†]http://www.bmrb.wisc.edu/

[‡]Triple-resonance pulse sequences utilise ^{1}H, ^{13}C and ^{15}N nuclei

HNCACB spectra show peaks at chemical shifts corresponding to both the C_α and C_β positions, respectively, for residues i and i-1. The CBCA(CO)NH spectrum shows C_α and C_β peaks for residue i-1 only, and therefore combined use of all three spectra reveals which chemical shifts correspond to which atom in which residue. Using the C_α and C_β shifts for i-1, a search is made for other amides in the HSQC that are correlated with the same shifts, and thus amides for residues i and i-1 are linked in sequence. The process is then repeated starting from residue i-1, eventually connecting up several residues.

In order to locate a group of connected residues in the known protein sequence, the pattern of C_α and C_β chemical shifts for a stretch of 3-4 residues is compared to a database containing characteristic shift data for each amino acid type. This usually generates a match which allows the connected residues to be assigned to a specific sub-sequence. C′ nuclear chemical shifts were assigned using the HNCO spectrum, which shows correlations between the amide of residue i and the carbonyl carbon (C′) of residue i-1. The $^1H^{15}N$-HSQC is shown in Figure 3.1. Backbone chemical shift data were used to predict secondary structure and backbone dihedral angles of *v*-Src SH2 (see Section 3.4.1 below).

3.3.2 Side chain assignment

The next stage in spectral assignment involved identification of chemical shifts for the side chain protons, carbons, and nitrogens. This was split into three phases corresponding to assignment of aliphatic groups, aromatic groups and amines. Prochiral aliphatic groups were not stereospecifically assigned.

3.3.2.1 Aliphatic resonances

Complete aliphatic side chain assignments were obtained by interpretation of spectra derived from through-bond and through-space experiments in the following manner:

Using the previously assigned $^1H^{15}N$-HSQC, backbone amide chemical shift coordinates were used to locate the cognate region of the ^{15}N-edited NOESY and TOWNY experiments. Protons at the α and β positions were identified by reference to HNHA and HNHB spectra, respectively, and their chemical shifts were correlated to previous assignment of C_α and C_β nuclei. Linkage between these two groups was confirmed independently using the HCCH-TOCSY spectra, which connects aliphatic protons within each residue side-chain. Unambiguous determination of α and β position shifts allowed assignment of the corresponding resonances in the $^1H^{13}C$-HSQC (Figure 3.2).

^1H and ^{13}C shifts for nuclei at the γ, δ, and ϵ positions were obtained where nec-

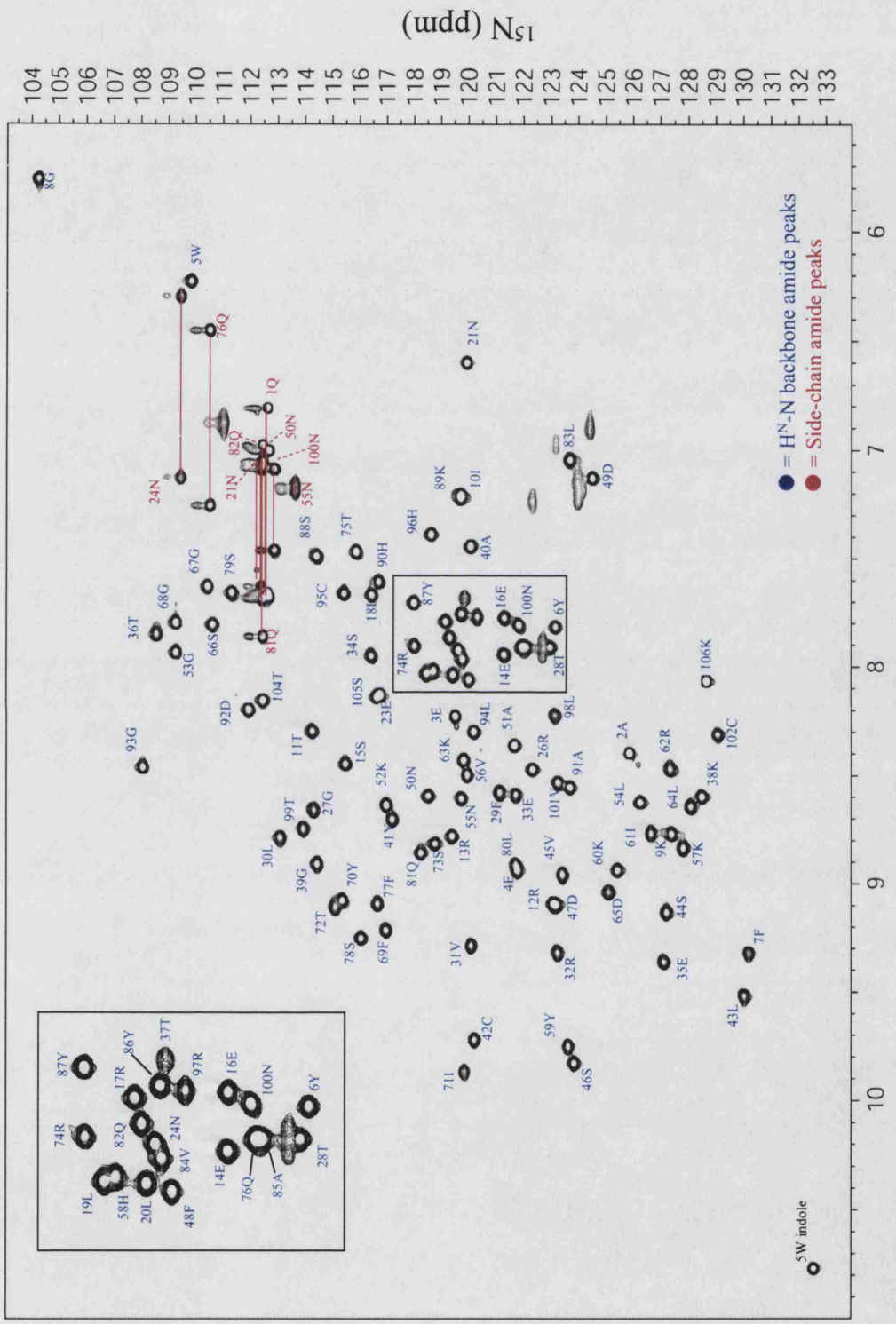


Figure 3.1 ^{1}H^{15}N-HSQC, assigned through analysis of triple-resonance spectra.

essary through further analysis of connectivities within the HCCH-TOCSY spectra, and confirmed with the triple-resonance experiment HCC(CO)NH, which correlates aliphatic carbon resonances (C_α to C_ϵ) for residue i-1 with backbone amide shifts for residue i.

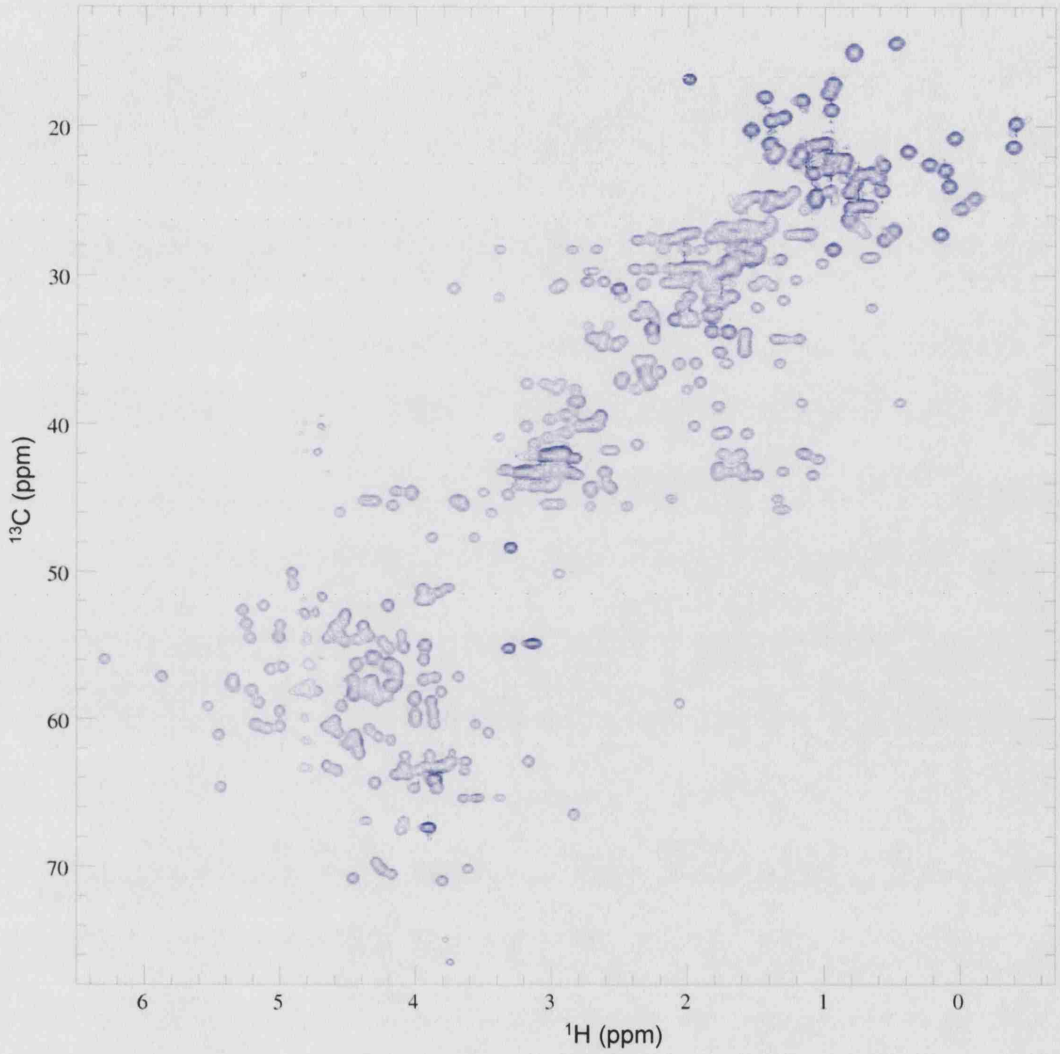


Figure 3.2 $^{1}H^{13}C$-HSQC constant time (600 MHz)

Due to the high quality of NMR data, assignment was straightforward for most protons, however certain spectral regions displaying signal degeneracy proved more difficult. These corresponded to arginine Hγ/Hδ and lysine Hδ/Hε protons. As these side chains are usually located on the surface of proteins, and therefore the electrochemical environment they experience is similar, this degeneracy is to be expected. One arginine (R32) could be assigned to non-degenerate resonances, and was found to be partially buried in the final structure. Suppression of signal around the water frequency (4.75 ppm) caused several resonances to be corrupted in the

^{13}C-edited spectra[§]. Their positions were deduced from the HCCH-TOCSY and triple-resonance spectra, yielding ^{1}H and ^{13}C shifts, respectively.

3.3.2.2 Aromatic resonances

Assignment of aromatic protons and carbons proved difficult as the HCCH-TOCSY and ^{13}C-edited NOESY spectra centred on the aromatic region were of poor quality[¶]. In the latter spectrum, very few inter-residue NOEs were observed, which impeded assignment. Consequently, the aromatic TOCSY was used to link individual resonances that belonged to the same ring and NOEs from aliphatic protons were used to assign connected peaks to a particular residue. Such assignments were confirmed by interpretation of the 2D (HB)CBHD experiment which correlates C_β and H_δ within the same aromatic side-chain (Figure 3.3). The assigned aromatic ^{1}H^{13}C-HSQC is shown in Figure 3.4

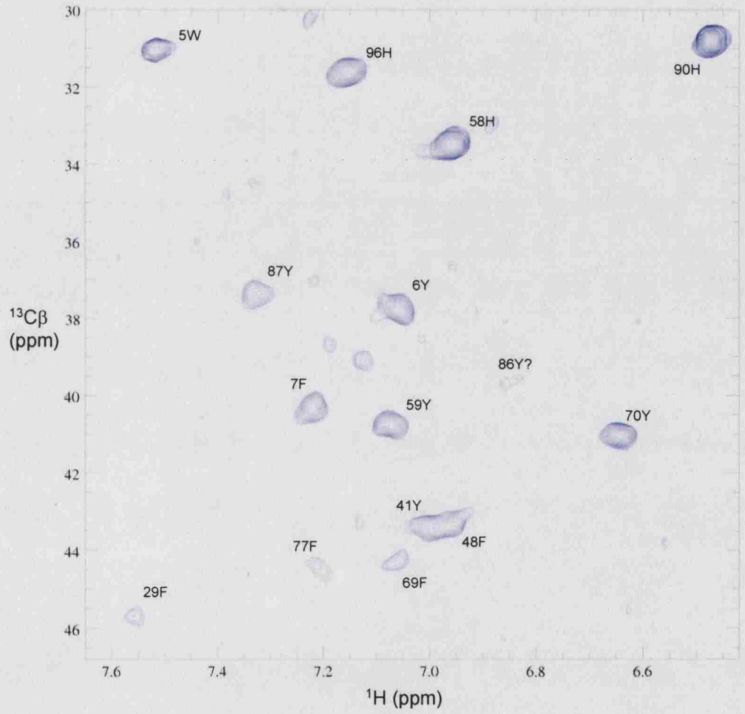


Figure 3.3 (HB)CBHD spectrum used to connect previously assigned $^{13}C\beta$ with $^{1}H\delta$ on aromatic rings.

Due to resonance overlap between the ring protons of the 5 phenylalanine residues, 4 ζ protons could not be located. Although the Hδ and Hε of F29 could be identified from NOESY and TOCSY spectra, no peak was observed in the ^{1}H^{13}C-HSQC. A similar situation existed for Y86 Hδ. Histidine resonances were simple to identify

[§]Six H_α protons (N21, F48, L54, N55, Y59, N100) and one Hβ (T11).
[¶]Pulse sequence not fully optimised at time of recording.

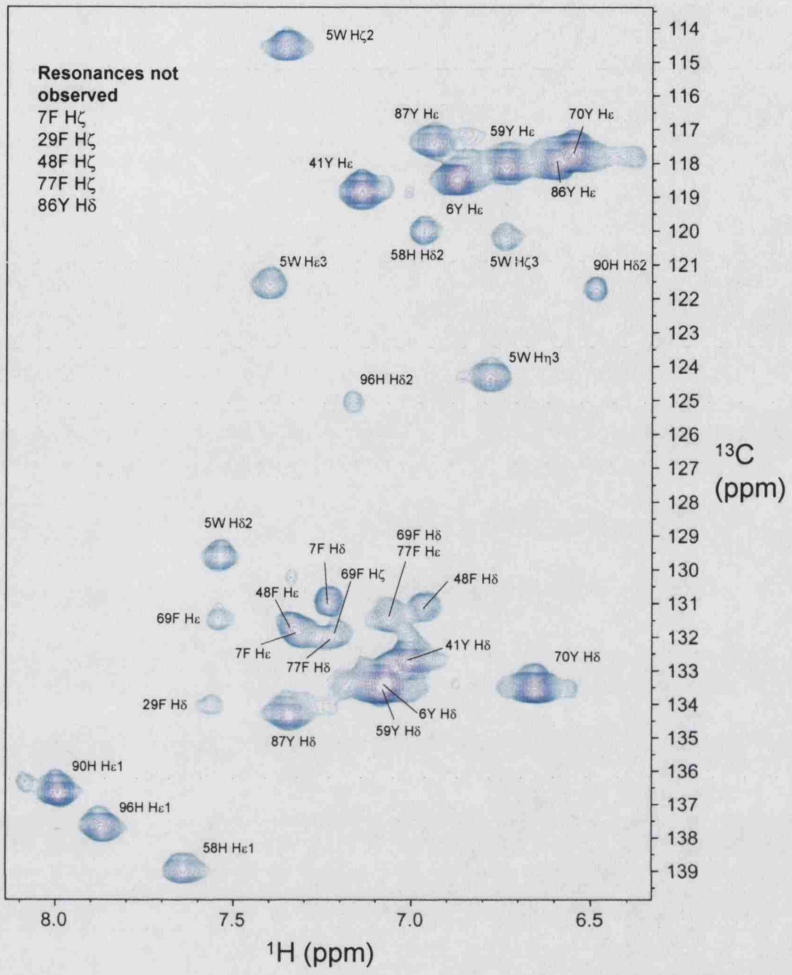


Figure 3.4 $^{1}H^{13}C$-HSQC centred on the aromatic proton region

and connect, however their assignment in sequence was not so obvious. Two of them were assigned on the basis of a single NOE each; the final histidine side chain completely lacked of NOE connections, and was therefore assigned by elimination.

3.3.2.3 Side chain amine resonances

Amines belonging to asparagine and glutamine side chains were identified using NOEs from both the ^{13}C- and ^{15}N-edited spectra. Hδ21 and Hδ22 were tentatively assigned on the basis of their relative chemical shift. The delocalised nature of the $-C(O)NH_2$ group results in slow rotation about the C–N bond and the distance between Hδ22 and the oxygen is shorter than for Hδ21 (2.30 and 3.08 Å, respectively[‖]). In free gln/asn, Hδ21 is less shielded than Hδ22, which renders the nucleus more susceptible to the local electromagnetic field and thus it displays a ~0.5 ppm

[‖]Measurements made using *PyMOL* (*DeLano Scientific*)

greater chemical shift than Hδ22.

In the context of a protein, hydrogen bonding or salt-bridge formation may alter or reverse this difference. It can be observed in the $^{1}H^{15}N$-HSQC spectrum that the amine protons of N55 have similar, yet distinct, chemical shifts – 0.033 ppm apart – whereas all the other amine proton pairs are separated by at least 0.37 ppm. This finding was explained on analysis of the final NMR structure, which indicated a network of hydrogen bonds (Figure 3.5). Since both amine protons of N55 are involved in hydrogen bonds, they experience similar electromagnetic environments, and hence display a smaller dispersion of chemical shift. It is possible in this case that Hδ21 actually has greater chemical shift of the two.

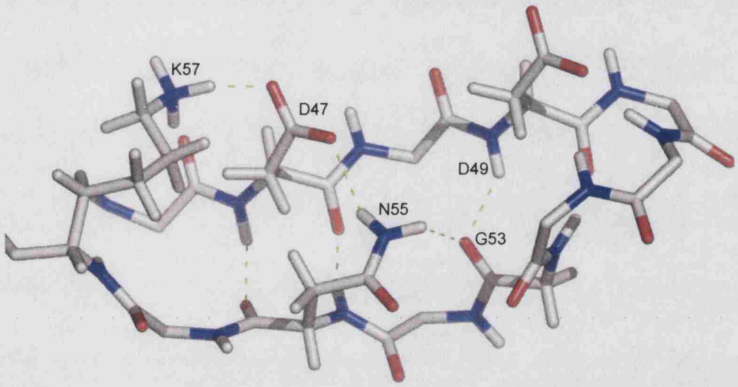


Figure 3.5 Hydrogen bond network involving side chain of N55

3.4 Generation of backbone torsion angle restraints

For peptides, steric hindrance prevents access to every possible dihedral angle, indeed only a relatively small range of values are observed in protein structures. Secondary structure is characterised by a regular pattern of ϕ/ψ dihedral combinations, commonly visualised in the Ramachandran plot. In order to guide formation of realistic secondary structure, these backbone angles are often restrained in structure calculations.

Chemical shifts alone can be used to predict protein secondary structure. Residues with α-helix and β-strand configurations display characteristic deviations in chemical shift from analogous random coil values. H_α and C_β nuclei have a higher chemical shift in β-strands and a lower shift in α-helices. C_α and C' nuclei display the inverse pattern.

The secondary structure of *v*-Src SH2 was investigated using three methods, all based on this fundamental observation, but of increasing sophistication. Firstly, deviations between observed chemical shifts and a standard library of random coil shifts

were calculated manually. Secondly, a consensus prediction of secondary structure was then made using *Chemical Shift Index* (*CSI*) software[237]. Finally, quantitative prediction of backbone angles was performed using *Torsion Angle Likelihood Obtained from Shift and sequence similarity* (*TALOS*)[70].

3.4.1 Manual comparison with random coil shifts

Residue-specific random coil chemical shifts** corresponding to H_α, C_α, C_β and C' nuclei were subtracted from observed values for *v*-Src SH2 and plotted (Figure 3.6). The characteristic pattern of C_α and C' shifts for residues 12-20 and 80-90 indicate the presence of two α-helices. Several β-strands, located at residues 29-33, 41-48, 53-60, 69-71, 75-77, are indicated by the plots, particularly in the C' and C_β data. No single nucleus unambiguously defines secondary structure, however the combination of results enables reasonable prediction at this early stage in the process of structure determination.

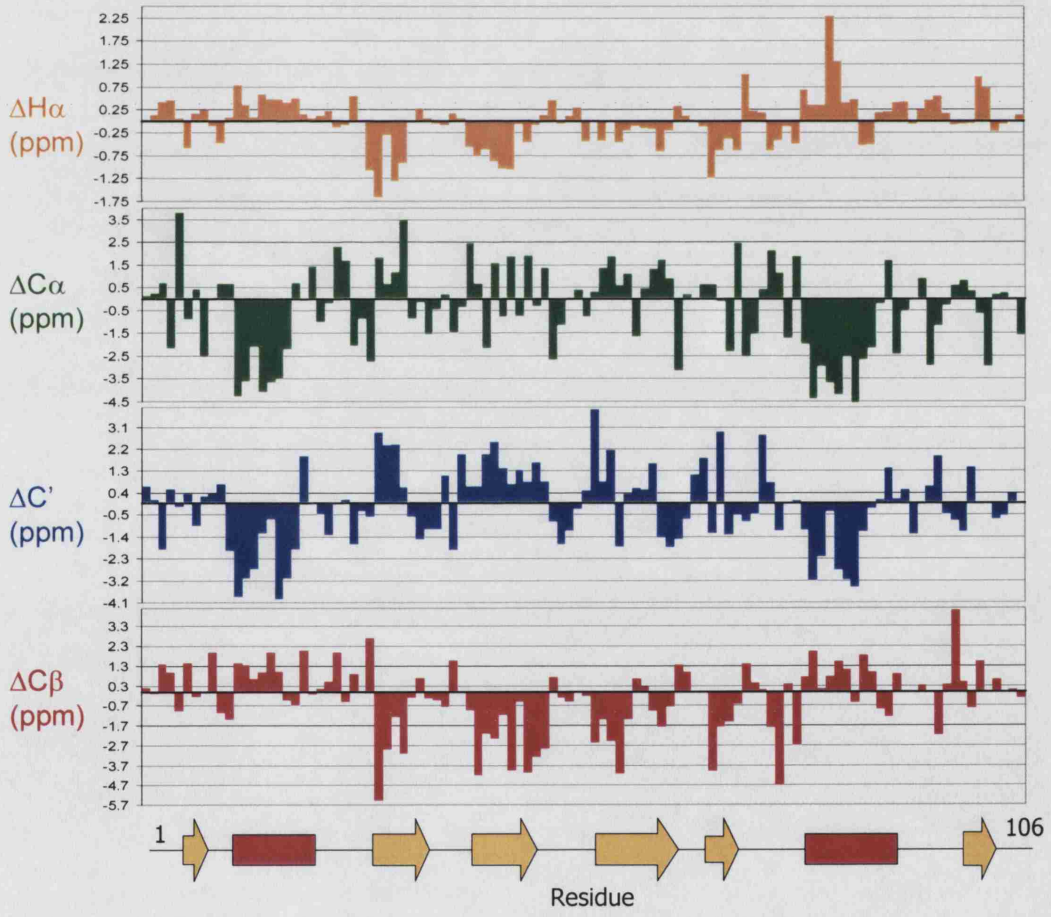


Figure 3.6 Deviation between observed chemical shifts and random coil values.

**Random coil shifts were derived from an ANSIG[226] library

3.4.2 *Chemical Shift Index*

The *CSI* program also compares observed H_α, C_α, C_β and C' chemical shifts to those contained in a library of random coil values, however this library was optimised empirically using a selected set of proteins[237]. Deviations in chemical shift are assigned a ternary chemical shift index value according to whether the random coil shift is smaller (-1), approximately equal to (0), or larger (+1) than the observed value. A consensus score is then derived for each residue by comparing indices for each nucleus.

H_α, C_α, C_β and C' shifts for *v*-Src SH2 were subjected to *CSI* analysis with the resultant prediction of three α-helices and five β-strands. The remainder of the sequence was classified as 'coil' - that is, of ambiguous structure. *CSI* predictions for the *v*-Src SH2 sequence are compared to observed secondary structure in both the NMR ensemble calculated in this study, and the peptide-free crystal structure (Figure 3.7).

	βA										αA													
CSI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
NMR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1SPR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

| | | | | | | | | | | | | | | | | | | | βB | | | | | | | | | | | | | | | | | | |

regions do not show distinctive shift differences to random coil values (Figure 3.6). *CSI* incorrectly predicts a short α-helix for residues 50-53, though a sharp turn in the polypeptide chain is found at this point.

Although no quantitative data from the two analyses described above was used in later structure calculations, knowledge of the location of secondary structure aided in manual NOE assignment. Since the predicted positioning of secondary structural elements matches closely that observed in the peptide-free/phosphate-bound crystal structure (1SPR[101]), these data indicate that the secondary structure of *v*-Src SH2 is the largely the same in the crystalline and solution states.

3.4.3 *TALOS* and dihedral angle restraints

The program *TALOS* attempts to convert backbone sequence and chemical shift patterns into dihedral angle information for use as geometric restraints in NMR structure determination projects[70]. Secondary structure clearly has a strong influence over backbone chemical shift patterns and a higher predictive accuracy can be achieved by considering the effect of amino acid sequence immediately surrounding the residue in question. Unlike the previous methods, *TALOS* does not compare observed chemical shifts to a random coil database. Instead, the program searches for sequence and shift matches within a database of proteins for which sequence, chemical shifts and high resolution structural data are available. The protein sequence in question is defined as a series of overlapping residue triplets and used to search a database of 9500 triplets from 78 proteins. Chemical shift data corresponding to matched sequences are then compared to the observed H_α, C_α, C_β, C', and N shifts, and the ten best shift matches reported. The ϕ/ψ angles relevant to the central residue in each database triplet are then analysed statistically, reporting the mean $\pm$ error.

Initially, angle predictions receive confidence flags (Good, Ambiguous, Bad) that reflect the variance within the matches. Results can then be examined and modified by hand, as necessary. Good predictions are reported to be 'accurate' in ~97 % of cases[††]. The small percentage of incorrect predictions can be identified though their inconsistency with NOE information or other supplementary data.

TALOS was used to analyse chemical shift and sequence data and suggest ϕ/ψ angles for the *v*-Src SH2 domain. Using the guidelines published for inspection and correction of predictions, 62 ϕ/ψ angle pairs (58 % of sequence) were classed as Good, 42 were Ambiguous, while N- and C-terminal residues received no prediction. No angular predictions were classed as Bad, however this is not unusual: 25 % of proteins analysed using *TALOS* result in no Bad predictions[70]. At an intermediate stage in

[††]Source: *TALOS* website, http://spin.niddk.nih.gov/bax/software/TALOS/index.html

the structural calculation process, the list of Good predictions was validated against a good quality NMR ensemble derived from NOE data alone (Figure 3.8). Only *TALOS*-derived predictions consistent with the structural ensemble were specified as restraints during further structure calculations.

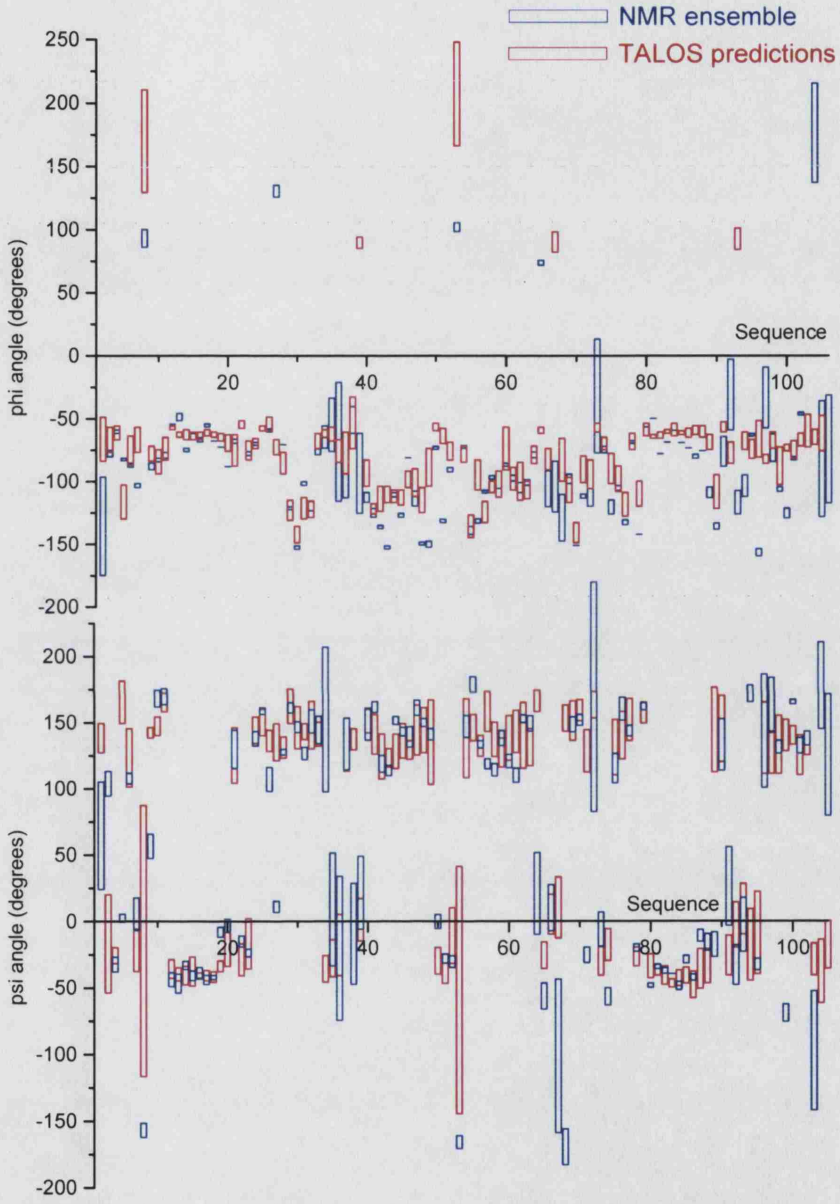


Figure 3.8 Comparison of predicted (red bars) and observed ensemble (blue bars) ϕ/ψ angle ranges. *TALOS* predictions are expressed as value $\pm$ error whilst the observed limits represent the mean $\pm$ standard deviation from the ensemble.

Through statistical analysis of these data, it was found that the predictive accuracy of *TALOS* was significantly lower than 97 %. 'Accuracy' could be defined in various ways – here, it was specified as the percentage of angle range predictions that include the actual ensemble-averaged mean value (see Figure 3.9). The accuracy of

ϕ and ψ predictions were determined for both Good and Ambiguous classes using angle ranges defined by σ and 2σ (Table 3.3). In this analysis, only 53 % of Good predictions were accurate when considering the $\theta \pm \sigma$ range. Surprisingly, 43 % of Ambiguous predictions were also accurate. Extending the predicted range to $\theta \pm 2\sigma$ increased the accuracy in both cases. Predictive accuracy for ϕ was consistently higher than for ψ, however the difference was usually less than 10 %.

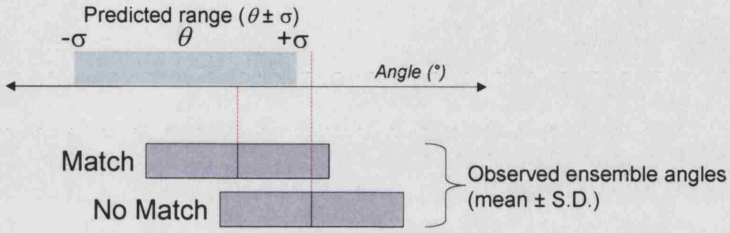


Figure 3.9 Assessing the accuracy of *TALOS* predictions. A hypothetical angle and error ($\theta \pm \sigma$) range predicted by *TALOS* is deemed to be accurate if the ensemble-averaged angle can be found within these limits.

Prediction class & range	ϕ (%)	ψ (%)
Good ($n = 62$)		
$\theta \pm \sigma$	53.0	50.0
$\theta \pm 2\sigma$	80.6	79.0
Ambiguous ($n = 42$)		
$\theta \pm \sigma$	42.9	38.1
$\theta \pm 2\sigma$	64.3	52.4
All ($n = 104$)		
$\theta \pm \sigma$	50.5	46.5
$\theta \pm 2\sigma$	76.2	70.3

Table 3.3 Statistical analysis of *TALOS* prediction accuracy. The percentage of ϕ and ψ angle predictions that were accurate is shown for the Good and Ambiguous classes, and the full data set. Angle ranges were defined as $\theta \pm \sigma$ and $\theta \pm 2\sigma$, where θ is the predicted angle and σ is the error.

In this case, *TALOS* results displayed significant inaccuracies – some confident predictions were incorrect, and some ambiguous predictions proved to be correct, depending on how stringently accuracy is defined. Angle restraints derived from *TALOS* should be treated with suspicion until they are verified through comparison with good quality ensembles derived from empirical data. The partial failure of *TALOS* to predict correct angles for *v*-Src SH2 may reflect an absence of appropriate sequence and shift information within the database. Also, a central assumption in the *TALOS* scheme is that backbone angles are equivalent in the solution and crystallographic state, allowing direct association of observed chemical shifts and dihedral angles. In some cases, this may not be valid.

The 33 'Good' predictions that were shown to be consistent with the ensemble were supplied as *CNS*-style angle restraints in subsequent structure calculations. Restraint ranges were specified according to the error in the prediction, or 15°, whichever was the greater.

3.4.4 Coupling constant restraints

Another useful class of angular restraints can be obtained through measurement of the magnitude of spin-spin couplings, or J-couplings. NMR peaks are often split by spin-spin coupling into multiplets whose frequencies are separated by an amount depending on the coupling constant, J. The magnitude of J is independent of the strength of the applied magnetic field and is given in units of frequency, Hz. Coupling constants depend on the electronic structure and hence on the relative orientation of the bonds connecting the atoms. Through the Karplus[238] relationship (Equation 3.1), coupling constants can be used as indirect restraints on particular protein angles.

$$^3J = Acos^2\theta + Bcos\theta + C \tag{3.1}$$

The constants A, B and C, that modulate this wave-like function have been defined empirically for a several θ dihedrals (e.g. ϕ,ψ,χ_1,χ_2). The coupling between H_N and H_α protons is dependent on ϕ, and is illustrated in Figure 3.10. As is evident from this curve, the same coupling constant can arise from multiple angles. From a structure determination point of view, this detracts from their usefulness as restraints. For example, a $^3J_{HNH\alpha}$ restraint of 6 ± 0.5 Hz would permit four different ϕ angle ranges ($-169.5 \ldots -161.5°$, $-78.4 \ldots -70.4°$, $30.1 \ldots 44.5°$, $75.5 \ldots 89.9°$).

The $^3J_{HNH\alpha}$ coupling constant is readily quantified through analysis of the HNHA spectrum[239]. In this scheme, the ratio of H_α and H_N peak volumes, denoted V_{H_α}/V_{H_N}, is related to magnitude of the J-coupling by the following equation:

$$^3J_{HNH\alpha} = \frac{\arctan\sqrt{Abs(V_{H_\alpha}/V_{H_N})}}{2\pi\zeta} \cdot CorrF \tag{3.2}$$

Observed H_N and H_α crosspeak volumes were extracted from the HNHA spectrum within ANSIG[226] and used to calculate per-residue $^3J_{HNH\alpha}$ coupling constants. The ζ and $CorrF$ experimental parameters were equal to 0.0134 and 1.11, respectively[239]. Since proline residues lack an H_N nucleus, no data was accessible for this residue. The 102 coupling constants (listed in Appendix B) were used as restraints in structure calculations. During this process, ϕ angles are back-calculated into coupling constants using Equation 3.1. If the value is outside of the error range (0.5 Hz) then an energetic penalty is applied.

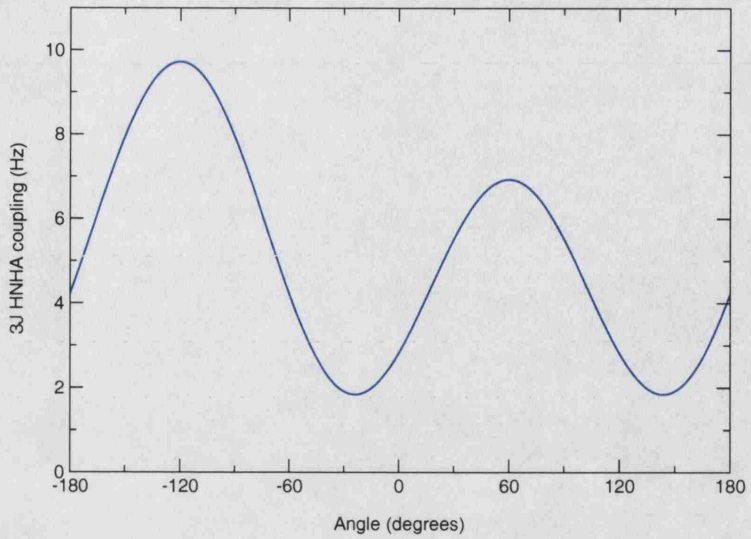


Figure 3.10 Karplus curve for HN-Hα. Data was simulated according to the empirically determined equation ${}^{3}J_{HNH\alpha}(\theta) = 6.4\cos^2(\theta - 60°) - 1.4\cos(\theta - 60°) + 1.9$ using *ORIGIN* (*OriginLabs*).

3.5 Hydrogen bond restraints

Hydrogen bonds make a significant contribution to protein structure, stability and ligand interactions (for a review see Williams & Ladbury[240]). The formation of such bonds is energetically favourable and therefore unsatisfied hydrogen bond donors are unusual[241]. Until recently, NMR pulse sequences could only provide indirect evidence for a particular hydrogen bond through observation of low amide proton solvent exchange rates or small temperature coefficients. Optimisation of the HNCO experiment allows detection of the very weak coupling across the N–H_N $\cdots$O=C hydrogen bond and even some degree of quantification of hydrogen bond lengths[242]. As the experiment is relatively insensitive, 6 days acquisition time on a $\geq$600 MHz spectrometer with protein concentrations in excess of 1 mM are required. Unfortunately, the *v*-Src SH2 domain was not stable under these conditions.

With the assumption that all hydrogen bond donors and acceptors are, in reality, satisfied one can attempt to locate such bonds on the basis of a known protein structure. Many protein hydrogen bonds involve backbone-backbone contacts, with interaction between backbone and side chains commonly found at helix caps[243, 244]. Hydrogen bonding patterns strongly define secondary structure conformation and thus restraining such interactions helps to define more realistic helical or β-sheet regions, improving the precision and accuracy of the structure. Hydrogen bonds within loop regions, or those involving side chains are harder to predict due conformational variability – there may be two or more possible acceptors. In this situation, NOE restraints are a better guide for structure calculations, though if a particular

hydrogen bond is consistently possible there is a stronger case for defining it.

The program *HBPLUS* searches for and analyses hydrogen bonds in PDB coordinate files[241]. The empirical definition of a hydrogen bond used by the program is based on statistical analysis of crystal structures, which revealed typical atomic geometries for hydrogen bond donor/acceptor pairs[245]. The most discriminating features are an H···A distance less than 2.5 Å and a D-H···A angle greater than 90°, where D and A represent donor and acceptor atoms, respectively. Using this threshold, the percentage of buried yet unsatisfied NH and CO groups for a dataset of 57 high-resolution crystal structures was 9.5 % and 5.8 %, respectively[241]. Increasing the threshold to 3.0 Å and >60° results in only 1.3 % (NH) and 1.8 % (CO) being unsatisfied.

HBPLUS was used to examine potential hydrogen bonding interactions in *v*-Src SH2 using a good quality NMR ensemble consisting of 20 water-refined structures*. Table 3.4 shows the average number of hydrogen bonds predicted per structure according to the four pairwise combinations involving either backbone or side chain donors and acceptors. For the backbone-backbone hydrogen bonds, average observed H···A and D···A distances were 2.04 ± 0.21 Å and 2.92 ± 0.19 Å, respectively. The average D-H···A angle was 150.3 ± 14.9°. These values agree closely with the most commonly observed distances and angles reported by McDonald & Thornton (see Figure 2 in ref[241]).

Donor	Acceptor	Mean ± SD
Backbone	Backbone	54.9 ± 2.0
Backbone	Sidechain	4.2 ± 1.7
Side chain	Backbone	11.9 ± 2.1
Side chain	Side chain	12.4 ± 3.6

Table 3.4 Hydrogen bond statistics observed for a good quality *v*-Src SH2 NMR ensemble

Predicted hydrogen bonds connecting backbone atoms compared to good quality NMR ensembles. Where a match occurred, appropriate distance restraints were supplied to constrain this interaction in subsequent structure calculations. The accepted hydrogen bonds are specified in Table 3.5. The optimal geometry of the hydrogen bond was obtained by specifying H···A and D···A restraint distances of 2.0 ± 0.3 Å and 3.0 ± 0.3 Å, respectively. In order to satisfy both these restraints, the D–H···A angle must be at least 103° (see Figure 3.11).

*Water-refined structures were used in this analysis since the force field treats non-bonded interactions, such as hydrogen bonds, in a more realistic manner

N-H_N	C'O	N-H_N	C'O
V31	W5	Y59	L43
F7	V31	K57	V45
E33	G8	D47	N55
E14	T11	N55	D47
S15	T11	D49	G53
E16	R12	G53	D49
R17	R13	R62	Y70
L18	E14	Y70	R62
L19	S15	L64	G68
L20	E16	F69	F77
T99	G27	F77	F69
S46	T28	Q82	S79
F29	N100	L83	S79
C102	F29	V84	L80
L30	S44	A85	Q81
S44	L30	Y86	Q82
R32	C42	Y87	L83
C42	R32	S88	V84
S34	A40	K89	A85
Y41	I61	H90	Y86
I61	Y41	L98	Y87
V45	K57		

Table 3.5 Hydrogen bonds connecting backbone atoms used in structure calculations

3.6 Generation of inter-proton distance restraints

Inter-proton distance restraints were obtained from analysis of ^{15}N and ^{13}C NOESY-HSQC spectra. Peak positions were detected automatically using *AZARA* (Wayne Boucher, University of Cambridge) and checked by hand for obvious errors. At first, NOESY spectra were assigned manually and the list of 2140 NOEs was used to generate preliminary structures that closely resembled the overall fold of the SH2 domain. An automated assignment and structure calculation method was then employed. This method was new to our group and the *v*-Src SH2 domain was thought to be an excellent test case for integrating the software into our NMR structure determination approach. The following section describes some of the concepts and challenges faced by automatic NOESY assignment.

3.6.1 Automatic NOE assignment

NOE assignment is a 'bottleneck' in NMR structure determination projects. This is because a large proportion of NOEs have multiple assignment possibilities, which in-

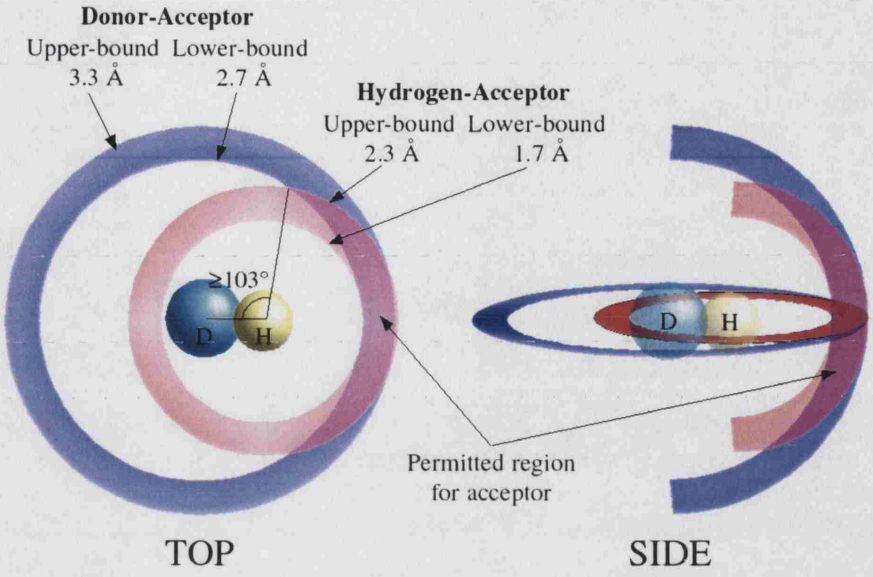


Figure 3.11 Upper and lower bound restraints on the H···A and D···A distances create optimal hydrogen bond geometry by indirectly specifying the D–H···A angle.

hibits the assembly of unambiguous restraint lists. This ambiguity can be gradually resolved by iterative cycles of NOESY assignment, structure calculation, and violation analysis. Several attempts have been made to automate this process, using algorithms to resolve ambiguity and discard inconsistent data[246, 247, 248, 249, 250, 251, 252]. One of the most widely-used programs is *ARIA* (*Ambiguous Restraints for Iterative Assignment*)[236, 249]. This software was used to obtain the solution structure of the apo *v*-Src SH2 domain. The following sections provide an overview of how *ARIA* handles NOESY assignment, structure calculations, and ensemble analysis.

A schematic diagram indicating the overall program flow of *ARIA* is shown in Figure 3.12. Each run begins with data initialisation, and is followed by 9 iterations (termed 0-8) of assignment and structural calculation.

3.6.1.1 Initialisation

Data validation

Chemical shift data and NOESY peaks are converted to XML format and stored in separate databases. The shift database classifies each atom according to one of three classes:

- Stereospecific – one atom corresponds to one unambiguous chemical shift;
- Equivalent – two or more equivalent atoms correspond to one degenerate chemical shift (e.g. methyl protons);

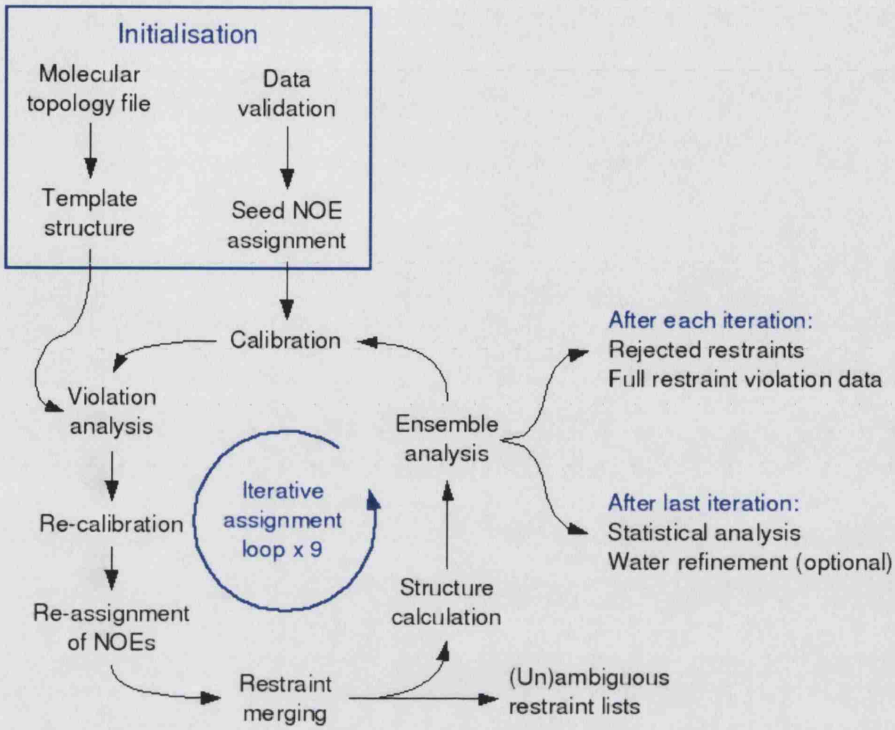


Figure 3.12 *ARIA* program flow

- Floating – two substituents of a prochiral group correspond to two different shifts (e.g. methylene protons). It is not known which atom corresponds to which shift.

Seed NOE assignment

The chemical shift database is used generate a list of possible assignments for every NOE. An NOE assignment is made if the chemical shift of a sequentially-assigned resonance falls within a user-specified distance of the position of the NOE peak.

Molecular system specification

The *CNS* script `generate.inp` is used to write a molecular topology file (MTF) that contains the name, chemical type, charge, and mass of each atom specified by the protein sequence, as well as the relevant covalent connectivities. Disulphide bridges and histidine protonation states can be easily specified; other modifications require further programming. The MTF is then used to generate the protein atomic coordinates in an extended conformation (*CNS* script `generate_extended.inp`) and write the coordinates in the PDB file format. Bond angles are randomised later and the structure is used as the initial template for assignment. The user can specify a folded structure instead, however assignment parameters must be adjusted accordingly to make use of this information during the run.

3.6.1.2 Iterative assignment and structure calculation

Each *ARIA* iteration consists of ensemble analysis, NOE calibration, removal of inconsistent NOEs, re-assignment of satisfied NOEs, merging of restraints and calculation of structures. These areas will be briefly discussed in turn.

Ensemble analysis

By default, 20 structures are calculated per iteration. These are ranked according to the total energy score (E_{tot}) and the seven lowest-energy conformers are used as a starting point for re-assignment in the next iteration.

NOE calibration

NOE peaks are quantified by their volumes or intensities, both of which are proportional to the inter-proton distance. The isolated spin pair approximation (ISPA) states that NOE volumes decrease with the 6^{th} power of distance:

$$V \propto d^{-6} \tag{3.3}$$

The actual NOE volume observed for a given inter-proton distance cannot be easily calculated as many natural processes (spin diffusion, internal dynamics, and chemical exchange) and experimental parameters affect the observed NOEs. An empirical calibration factor, α is therefore introduced to modulate the relationship:

$$V = \alpha d^{-6} \tag{3.4}$$

For iterations 1-8, the ensemble of structures calculated in the previous iteration is used to estimate the value of α by calculating the ratio of the average experimental volume V^{exp} to the average theoretical volume V^{theo}:

$$\alpha = \frac{\sum_{i=1}^{n} V_i^{exp}}{\sum_{i=1}^{n} V_i^{theo}} \tag{3.5}$$

V^{theo} is back-calculated from the average distance between pairs of protons using either the ISPA method (3.4), or relaxation matrix analysis. Iteration 0 is special since no structure ensemble exists for which to compare observed and back-calculated volumes. Instead, the calibration factor (α_0) is estimated using another constant, d_{avg}, which corresponds to the average distance between spin pairs causing NOEs:

$$\alpha_0 = n^{-1} \sum_{i=1}^{n} \frac{V_i^{exp}}{d_{avg}^{-6}} \tag{3.6}$$

The calibration factor α (or α_0) is then used to calculate a distance target for each observed NOE (Eq. 3.7). This distance is an estimate and it is common practice to

define lower (3.8) and upper (3.9) bound restraints instead:

$$D_i = (\alpha^{-1}V_i)^{-1/6} \quad (3.7)$$

$$Lower = max(0, D_i - \Delta) \quad (3.8)$$

$$Upper = D_i + \Delta \quad (3.9)$$

$$\text{where } \Delta = 0.125(D_i)^2 \quad (3.10)$$

The parameter Δ was determined empirically in such a way that it allows for typical cumulative errors due to spin diffusion, internal dynamics and peak integration[236]. For estimated distances of 2 and 4 Å, Δ is equal to 0.5 and 2 Å, respectively.

So far, only the ideal case of an unambiguous single-spin – single-spin NOE has been described. NOEs involving two or more equivalent spins can be treated in a similar manner. Let I and J represent two groups of equivalent spins with n_I and n_J members (e.g. two methyl groups). Although one peak is observed, it contains summed contributions from $n_I * n_J$ NOEs. Therefore, Eq. 3.4 is modified such that the distance used is the average pairwise distance, $\hat{d}_{IJ}$:

$$\hat{d}_{IJ} = (n_I n_J)^{-1} \sum d_{IJ} \quad (3.11)$$

$$\text{and } V_{IJ} = \alpha n_I n_J \hat{d}_{IJ}^{-6} \quad (3.12)$$

Ambiguous distance restraints

Limited dispersion of chemical shifts leads to degeneracy of resonances with the result that a large fraction of NOEs can only be assigned ambiguously. Without further information it is impossible to distinguish between spins that are involved in the NOE and those far apart, but which match the chemical shift. These NOEs clearly cannot be converted into meaningful distance restraints in the manner described above. It is of course also possible that a NOESY peak is actually the summed contributions from two or more individual NOEs, in this case the relative weight of each would be unknown.

The standard method of countering this ambiguity is to calculate preliminary structures using only unambiguous restraints. These preliminary structures often have the correct overall fold of the protein, which allows resolution of some ambiguities. An iterative procedure of reassignment and structure calculation is followed, gradually increasing the number of unambiguous assignments. One of the most important concepts in *ARIA* is that ambiguous information can be combined to give unambiguous results. This allows ambiguous NOEs to be converted into *ambigu-*

ous distance restraints (ADRs). This is defined as an effective distance ($\overline{D}$) that contains contributions from each pair of protons temporarily assigned to an NOE. For example, consider an NOE between an atom (A) and one of two assignment possibilities (B and C). The effective NOE volume V^{eff} is proportional to the sum of inverse sixth powers of the individual proton distances (3.13), which corresponds to an effective summed distance (3.14):

$$V^{eff} \propto d_{AB}^{-6} + d_{AC}^{-6} \tag{3.13}$$

$$\overline{D} = \left(d_{AB}^{-6} + d_{AC}^{-6}\right)^{-1/6} \tag{3.14}$$

$\overline{D}$ is then converted to lower- and upper-bound restraints as before (3.8-3.9). During the simulated annealing (SA) process, the distances A–B and A–C are back-calculated into volumes and summed.

Detection of inconsistent peaks

One of the tasks of automatic assignment is to identify and exclude crosspeaks in the NOESY spectrum that are inconsistent with the overall structure. This occurs when an NOE has been erroneously picked (e.g. random noise, NOEs to water), or where the position of the crosspeak is incorrect. Distance restraints arising from these inappropriate crosspeaks are expected to be inconsistent with each other and with the true molecular structure provided that their influence is largely outweighed by correct NOE data. A simple violation analysis is performed to identify inconsistent peaks. The fraction R_{vio} of structures in which a particular restraint is violated by more than a generous threshold v_{tol} is calculated. If this fraction exceeds a user-defined tolerance R_{tol} then the restraint is altered or discarded.

$$\text{If } R_{vio} > R_{tol} \text{ then discard restraint} \tag{3.15}$$

$$\text{where } R_{vio} = \frac{\sum[\Theta(L_i - d_i + v_{tol}) + \Theta(d_i - U_i - v_{tol})]}{S} \tag{3.16}$$

where d_i is the observed inter-proton distance for NOE i from S converged structures, U_i is the upper- and L_i is the lower-bound restraints, respectively. $\Theta(x)$ is the Heaviside step function, which takes a value of 1 if $x > 0$, otherwise 0. R_{tol} is normally held constant at 0.5 throughout the 9 iterations; v_{tol} is gradually reduced from 1000 to 0.1 Å.

To display the operation of this violation analysis, the following example is given. Let the distance between two nuclei, A and B, be restrained by lower- (L_{AB}) and upper-bound (U_{AB}) limits of 2 and 5 Å, respectively. After simulated annealing,

the 7 lowest energy structures are identified, the distances between A and B are measured (d_{AB}), and compared to the original restraint limits (Table 3.6). In five of the seven structures (i.e. $R_{vio} = 0.71$), $d_{AB} > U_{AB} + v_{tol}$, and therefore the threshold set by R_{tol} is exceeded and the restraint is not used in further iterations. After violation analysis, the remaining NOE data is re-calibrated using Equation 3.5.

d_{AB} (Å)	$d_i - U_i - v_{tol}$	$\Theta(d_{AB} - U_{AB} - v_{tol})$	Violated?
5.3	-1.7	0	No
5.8	-1.2	0	No
7.2	0.2	1	Yes
7.5	0.5	1	Yes
8.9	1.9	1	Yes
9.7	2.7	1	Yes
12.0	5.0	1	Yes
	Total:	$5/7 > R_{tol}$	

Table 3.6 Example of a consistently violated restraint. The values used for v_{tol} and R_{tol} were 2 Å and 0.5, respectively.

Partial assignment

Each *ARIA* iteration involves reassignment of the NOESY spectrum, with the aim of eliminating unlikely contributions to each NOE. For a given ambiguous restraint, individual assignment possibilities are quantified by back-calculating the expected volumes corresponding to the observed ensemble-averaged distances:

$$C_i = \hat{d}_i^{-6} \tag{3.17}$$

The total contributions to an NOE are normalised such that they sum to 1.0 and are ranked, largest first. The number of assignment possibilities is then reduced by keeping the x largest contributions that satisfy

$$\sum_{1}^{x} C_i \geq p \tag{3.18}$$

where p is a user-defined ambiguity cutoff. This parameter is decreased after each *ARIA* iteration causing more restraints to be fully assigned in the normal sense of the word. In the first iteration, p is set to 1, effectively allowing all possible assignments to proceed into the simulated annealing stage. The following example represents a typical case in which the level of ambiguity is reduced for a particular assignment:

Proton x displays an NOE that has up to four potential assignments, i, j, k, and l. The average distances between x and each partner are calculated from the previous structure ensemble and their relative contributions to the NOE calculated using

Equation 3.17 (see Table 3.7). The ambiguity cutoff p was set at 0.9. Contribution i alone does not exceed this value, but by summing the top two contributions, Equation 3.18 is satisfied. Consequently, assignment possibilities k and l are excluded from further calculations.

Proton pair	Contribution	Cumulative total
i	0.60	0.60
j	0.32	0.92
k	0.06	0.98
l	0.02	1.00

Table 3.7 Hypothetical contributions to an ambiguous NOE

Exclusion of very weakly contributing assignments is advantageous for two reasons. Firstly, they are likely to correspond to incorrect assignments and relieving their association with a particular proton allows them to satisfy more easily NOEs in which they are significant contributions. Secondly, even if they are small yet real contributions to the NOE, they would have little effect on the final structure. From the point of view of throughput, a reduction in data ambiguity is leads to a decrease in calculation time. One can limit in each iteration the maximum number of assignment possibilities allowed (n_{max}). For moderately-sized proteins, a value of 20 is suggested[253].

Merging

Duplication of restraints within and between NOESY datasets leads to bias at those positions. As a rule, a restraint is discarded if an equivalent version is found elsewhere with shorter distance bounds.

Structure calculation

ARIA interfaces directly to *CNS* and employs simulated annealing protocols for structure calculations. Conditions are adapted to ensure proper treatment of ambiguous distance restraints. Details concerning simulated annealing methods can be found in Section 3.6.2.3.

Floating chirality assignment

As stereospecific assignment of methylene protons and isopropyl groups is often difficult to achieve, distance restraints involving such ambiguities require special attention. *ARIA/CNS* utilises a floating assignment method in which stereospecific assignment possibilities are randomly swapped during calculations. The energetically-preferred solution is written to a data file.

Water refinement

The simulated annealing protocol used during iterative assignment involves simplified non-bonded force potentials in a dehydrated state in order to keep calculation times to a minimum. Consequently, structures display artefacts such as unrealistic side chain packing and unsatisfied hydrogen bond acceptors and donors[236]. To improve structural quality, *ARIA* performs solvent refinement on the final ensemble of converged structures. This involves calculating a short trajectory with a full molecular dynamics force field, which includes electrostatic and Lennard-Jones potentials. Where relevant, parameters used in refinement are the same as those used for simulated annealing to avoid systematic differences.

3.6.2 Use of *ARIA* for the solution structure of *v*-Src SH2

ARIA v2.0a was used for iterative assignment of NMR data and structure calculations with the *v*-Src SH2 domain. The software assumes that the NOE crosspeak lists correspond to genuine intra-protein NOEs. As described above, NOESY peaks were automatically picked using *AZARA*, and edited carefully by hand to remove obvious errors. However, the data-set remained imperfect and incomplete, leading to highly improbable distance restraints, which due to their inconsistency with correct data, and with each other, were subsequently violated in structure calculations. NOE crosspeaks at the water frequency were commonly mis-assigned to several H_α resonances. These were gradually identified and removed through successive *ARIA* runs.

Another class of erroneous restraints originated from automatic peak picking inadequacies. Two NOEs that are almost degenerate in chemical shift occur as oval-shaped peaks in the spectrum. *AZARA* was not able to distinguish this situation from that of a single, isolated NOE, and therefore placed a crosspeak at the geometric summit, which occurs at an intermediate chemical shift between the two real NOEs. When *ARIA* attempts to assign such an NOE, it misses the correct candidate(s), and suggests incorrect alternatives instead. This is highlighted by violation analysis and solved by manual placement of two crosspeaks at the correct coordinates, followed by re-integration of the spectrum. Sometimes, this type of error led to restraints that were almost consistent with the overall data, and thus evaded detection until later stages of refinement.

ARIA was used to assign NOEs derived from four experiments – two ^{15}N-edited NOESY-HSQC spectra with mixing times of 200 and 50 ms, and two ^{13}C-edited NOESY-HSQC spectra (both 125 ms mixing time), one of which covered the aromatic region. The two ^{15}N NOESY-HSQC data sets were of high quality as they had been previously assigned to a high degree of completion and most errors had been

removed prior to use with *ARIA*. The ^{13}C NOESY-HSQC data was of poorer quality due to relatively high background noise levels and unsolvable problems with intense resonances such as methyl groups*. ^{13}C-edited spectra were only partially assigned by hand; it thus contained many crosspeak errors on introduction to *ARIA*. Tolerance windows around each NOESY peak were set to 0.028 × 0.030 × 0.30 ppm and 0.032 × 0.032 × 0.30 ppm for ^{15}N- and ^{13}C-NOESY spectra, respectively. The first and second values refer to the two proton dimensions; the third value corresponds to the heteronuclear dimension.

Initial usage of *ARIA* relied exclusively on empirical NOE intensities data to generate a reliable ensemble. Coupling constants were introduced later to restrain ϕ backbone angles. Individual *TALOS*-derived ϕ/ψ angle predictions were included later, but only if they were consistent with the structure ensemble, a situation that was constantly monitored through violation analysis. Finally, hydrogen bond restraints were supplied in order to improve the quality of secondary structure conformation. *PROCHECK-NMR* software was employed to provide structural quality indices and assist in the identification of regions that suffer a high degree of violation[254]. *ARIA* also writes restraint files in *MOLMOL* format, which can be used to observe the location of restraints and their degree of violation in three dimensions[255].

3.6.2.1 Floating chirality

The protons in methylene and isopropyl groups were not assigned in a stereospecific manner in the original chemical shift lists. Consequently, these protons were treated using the floating chirality method during structure calculations. Stereospecific assignments are suggested on the basis of each structure, however this information is not used during subsequent iterations or structure calculations – its only purpose is to aid stereospecific assignment in the final structure.

It was discovered on cursory examination of the input XML data that isoleucine $H_{\gamma 2}$ and $H_{\delta 1}$ methyl proton chemical shifts were being treated as 'floating', despite the fact that they were stereospecifically assigned†. This increases the complexity of assignment and structure calculations unnecessarily, and may have a negative impact on structural convergence and quality. Therefore the XML code describing each isoleucine was altered as shown in Figure 3.13. The isoleucine methyl groups were separated in the code, and given the appropriate single chemical shift that

*Unfortunately, technical problems were being experienced at the NIMR 800 MHz facility at the time of recording. Intense diagonal (self-NOE) peaks displayed linear extensions of signal intensity that encompassed nearby off-diagonal peaks. This led to a significant artefactual enhancement of peak intensity for these NOEs, which prevented them being used in structure calculations.

†This bug has now been fixed by Dr Mark Williams.

describes the equivalent protons of an isolated methyl group. A similar situation existed for stereospecifically assigned asparagine and glutamine amine protons. This was also fixed by adapting the XML code (Figure 3.13).

Isoleucine methyls treated as 'floating'

```
<shift_assignment method="FLOATING">
   <spin_system averaging_method="FAST">
      <atom segid="    " residue="10" name="HD11"/>
      <atom segid="    " residue="10" name="HD12"/>
      <atom segid="    " residue="10" name="HD13"/>
      <chemical_shift value="0.788" error=""/>
      <chemical_shift value="0.972" error=""/>
   </spin_system>
   <spin_system averaging_method="FAST">
      <atom segid="    " residue="10" name="HG21"/>
      <atom segid="    " residue="10" name="HG22"/>
      <atom segid="    " residue="10" name="HG23"/>
      <chemical_shift value="0.788" error=""/>
      <chemical_shift value="0.972" error=""/>
   </spin_system>
</shift_assignment>
```

Isoleucine methyls treated as stereospecific

```
<shift_assignment method="EQUIVALENT">
   <spin_system averaging_method="FAST">
      <atom segid="    " residue="10" name="HD11"/>
      <atom segid="    " residue="10" name="HD12"/>
      <atom segid="    " residue="10" name="HD13"/>
      <chemical_shift value="0.788" error=""/>
   </spin_system>
</shift_assignment>

<shift_assignment method="EQUIVALENT">
   <spin_system averaging_method="FAST">
      <atom segid="    " residue="10" name="HG21"/>
      <atom segid="    " residue="10" name="HG22"/>
      <atom segid="    " residue="10" name="HG23"/>
      <chemical_shift value="0.972" error=""/>
   </spin_system>
</shift_assignment>
```

Asn/Gln amine protons treated as 'floating'

```
<shift_assignment method="FLOATING">
   <spin_system averaging_method="NONE">
      <atom segid="    " residue="1" name="HE21"/>
      <chemical_shift value="6.80683" error=""/>
      <chemical_shift value="7.65769" error=""/>
   </spin_system>
   <spin_system averaging_method="NONE">
      <atom segid="    " residue="1" name="HE22"/>
      <chemical_shift value="6.80683" error=""/>
      <chemical_shift value="7.65769" error=""/>
   </spin_system>
</shift_assignment>
```

Asn/Gln amine protons treated as stereospecific

```
<shift_assignment method="STEREOSPECIFIC">
   <spin_system averaging_method="NONE">
      <atom segid="    " residue="1" name="HE21"/>
      <chemical_shift value="6.80683" error=""/>
   </spin_system>
</shift_assignment>

<shift_assignment method="STEREOSPECIFIC">
   <spin_system averaging_method="NONE">
      <atom segid="    " residue="1" name="HE22"/>
      <chemical_shift value="7.65769" error=""/>
   </spin_system>
</shift_assignment>
```

Figure 3.13 Changes made to the chemical shift database (XML-format) for stereospecifically-assigned ile (*top*) and asn/gln (*bottom*) side chains, which are wrongly being treated as 'floating'. The original code is shown on the left and the fixed code on the right. Changes to the code are highlighted in red colour.

3.6.2.2 Histidine protonation states

Chapter 4 describes NMR-derived pK_a values for titratable groups in *v*-Src SH2. It was found that H58 (His βD4) is completely deprotonated at the $N_{\delta 1}$ position at pH 6.0, and that H96 is two-thirds deprotonated. H90 has a pK_a of 6.4, and is approximately one-third deprotonated at pH 6.0. These findings are consistent with previous studies of the SH2 domain of PLCγ[256]. Therefore, during structure calculations the $H_{\delta 1}$ proton was removed from H58 and H96 in order make the simulated annealing process more realistic. This was achieved by activating the 'HISE' patch for H58 and H96 within the *ARIA* graphical user interface (GUI).

3.6.2.3 Run parameters and simulated annealing conditions

Default *ARIA* v2.0a run parameters (Table 3.8) were used to generate preliminary structures and help identify errors in the NOESY peak lists. After repeated usage of the program (estimated at about 75 applications!), a higher quality data set was

obtained. Subsequent runs involved calculation of a larger number of structures for iterations 5-8 in order to increase the number of low-energy structures from which to select a representative ensemble. The modified run parameters are shown in Table 3.9.

	ITERATION NUMBER								
	0	1	2	3	4	5	6	7	8
No. structures calculated	20	20	20	20	20	20	20	20	20
Use best n structures	7	7	7	7	7	7	7	7	7
Violation tolerance (v_{tol} Å)	1000	5	3	1	1	1	0.1	0.1	0.1
Violation threshold (R_{tol})	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Ambiguity cut-off (p)	1.0	0.9999	0.999	0.99	0.98	0.96	0.93	0.9	0.8
Max. no. contributions (N_{max})	20	20	20	20	20	20	20	20	20

Table 3.8 Standard *ARIA* run parameters governing initial iterative assignment

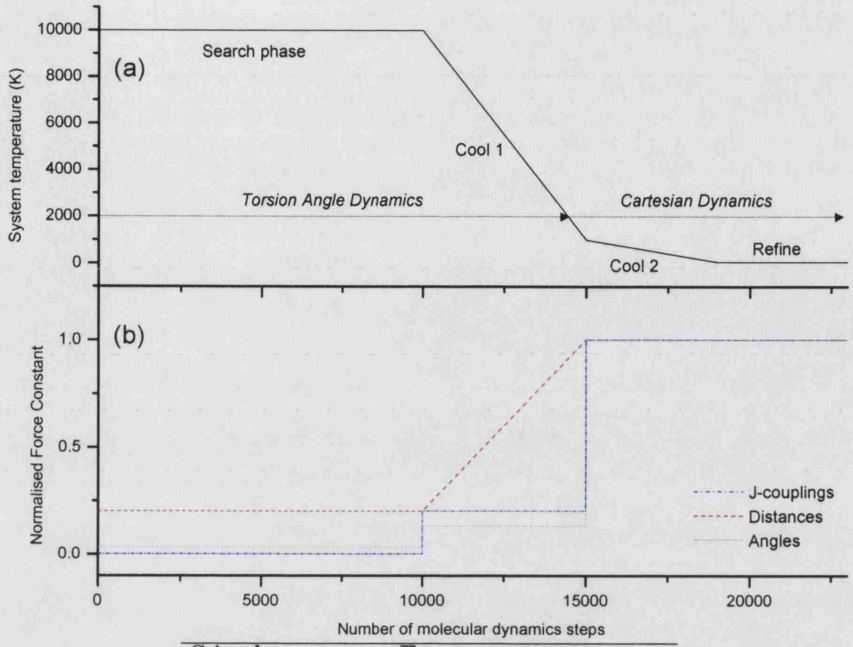
	ITERATION NUMBER								
	0	1	2	3	4	5	6	7	8
No. structures calculated	20	20	20	20	20	40	60	60	200
Use best n structures	10	10	10	10	10	10	20	20	100

Table 3.9 Modifications to *ARIA* parameters for late stage iterative assignment

Structures were calculated by *ARIA/CNS* using simulated annealing protocols in which the molecule is heated and then cooled slowly whilst the target function (energy) is subjected to minimisation[257]. Covalent geometry was described by `parallhdg5.3`[258] and the non-bonded part of the force field was defined by the PROLSQ parameter set[259]. Violations of known geometries and experimental restraints suffer energetic penalties according to a flat-bottom harmonic wall function. The SA protocol consists of four phases throughout which the system temperature is regulated – search, cool 1, cool 2, and refine (Figure 3.14a). During the high-temperature search phase, violations are tolerated in order to allow efficient sampling of a large number of possible structures. Rapid variation in atomic positions is achieved through bond rotation using a torsion angle dynamics (TAD) algorithm[260]. The cooling phases (10000 → 1000 → 50 K) feature a gradual increase in the weighting of contributions from the force field and experimental restraints (Figure 3.14b). Below 2000 K, Cartesian dynamics algorithm is used to vary atomic positions instead of TAD. The last phase (refine) permits further minimisation at the target temperature. The simulated annealing parameters used to calculate the structure of *v*-Src SH2 were unchanged from the default values supplied with *ARIA* v2.0a.

The 100 lowest-energy[‡] structures from iteration 8 were refined further in a thin layer of explicit solvent (H_2O). Default parameters were used to control the heating, refinement, cooling and final minimisation phases. The 100 water-refined structures were analysed using a combination of *ARIA* scripts, *PROCHECK*, and visual

[‡]Structures were ranked according to total energy (E_{tot}) rather than NOE restraint energy (E_{noe}) in order to include contributions from other restraints and the force field.



SA phase	Force constants		
	Distances	Angles	J-couplings
Search	10	5	0.0
Cool1	10 → 50	25	0.2
Cool2	50	200	1.0
Refine	50	200	1.0

Figure 3.14 Control of system temperature (a) and force constants (b) during simulated annealing. The force constants are normalised to 1.0 in the figure; actual values are displayed in the accompanying table.

comparison. Conformers that displayed significant violations (i.e. large E_{tot} and/or large E_{noe}), or poor structural quality indices were discarded from the group. The remaining 22 structures were deposited in the Protein Data Bank under accession code 1Z9R[§].

3.7 Analysis of *v*-Src SH2 solution structure

The accepted ensemble of solution state *v*-Src SH2 structures is shown in Figure 3.15, along with a representative (lowest-energy) structure. A summary of ensemble-averaged structural and restraint statistics can be found in Tables 3.10 and 3.11, respectively. The calculated structures display the characteristic of the SH2 domain family fold: two parallel α-helices surrounding a central, four-stranded, anti-parallel β-sheet, contiguous with a second, smaller β-sheet consisting of three short, anti-parallel strands. The pY and pY+3 pockets are also evident within the ensemble, as is the hydrogen bonding network that serves to orientate the crucial Arg βB5

[§]The entry will be made available to public access with an accompanying publication that is currently in preparation.

(see Chapter 4). Side chains involved in this network (E16, S44, and H58) show relatively low ensemble-averaged heavy atom rms deviations from the mean (~0.7 Å) compared with other important residues for binding, such as R12 (2.2 Å), S34 (1.0 Å) and K60 (1.49 Å).

Parameter	Mean	±	SD
Structural precision[a] (Å)			
Backbone atoms, secondary structure	0.248	±	0.036
Heavy atoms, secondary structure	0.623	±	0.063
Backbone atoms, all residues	0.775	±	0.087
Heavy atoms, all residues	1.218	±	0.097
Deviations from idealised covalent geometry			
Bond lengths (Å)	0.0066	±	0.0003
Bond angles (°)	0.819	±	0.024
Impropers[b] (°)	2.475	±	0.078
Energies (kcal mol^{-1})			
E_{total}	-3150.92	±	111.23
E_{bonds}	75.79	±	7.54
E_{angles}	317.64	±	17.73
$E_{impropers}$	854.71	±	54.72
$E_{dihedrals}$	567.35	±	9.14
E_{vdw}	-966.82	±	15.96
$E_{electrostatics}$	-3999.6	±	95.0
E_{noe}	153.48	±	34.91
Structural quality (*PROCHECK*)[c]			
Most favoured (%)	83.3	±	2.8
Additionally allowed (%)	14.0	±	2.2
Generously allowed (%)	1.2	±	1.1
Disallowed (%)	1.6	±	0.6
No. bad contacts[d]	0.1	±	0.3

[a]Precision of atomic coordinates is defined as the average pairwise RMSD between each of the 22 conformers and the mean coordinate structure over all backbone residues
[b]Improper torsion angle restraints maintain planarity and chirality where appropriate
[c]Stereochemical parameters were analysed using *PROCHECK*. The values correspond to the percentage of residues with backbone phi and psi values that fall within defined regions of the Ramachandran plot.
[d]This value is normally between 0 and 30 for a crystal structure with resolution greater than 3.0 Å

Table 3.10 Summary of structure statistics

Statistics concerning the precision and violation of each restraint type are shown in Table 3.11. The results show that the structures match closely the distance and angular restraints that are supplied at the start of the simulated annealing process. It is important to remember that structural convergence is strongly influenced by the nature of the force field, which towards the end of the run imposes a large energetic penalty on violations. Therefore, as long as the data is largely self-consistent, a low rms deviation will be achieved. In other words, a few serious violations within an otherwise sound data set will have a negligible effect on the rms deviation since the large quantity of satisfied restraints will outweigh such contributions. Thus it is useful to consider the number and location of violations within the ensemble.

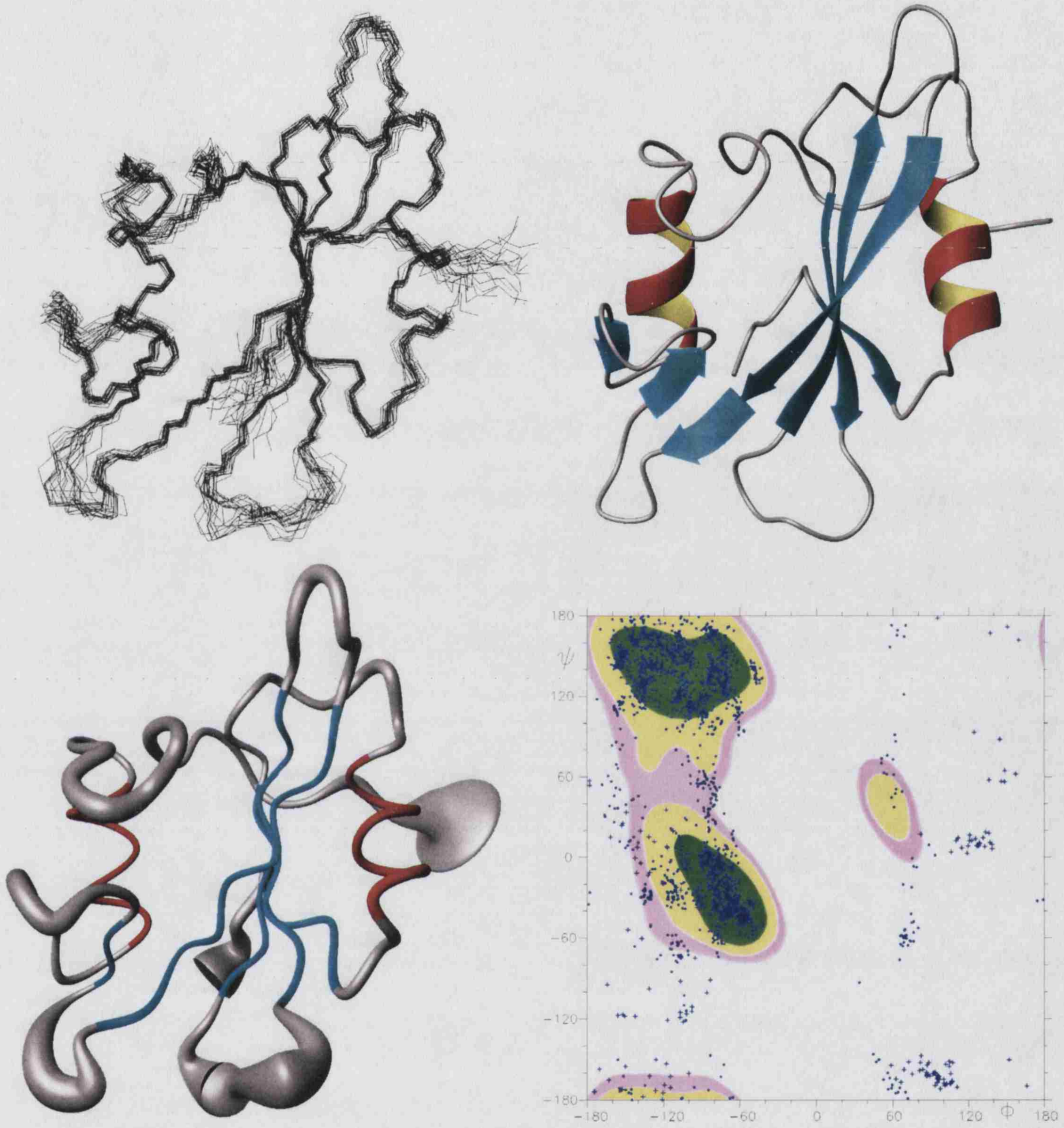


Figure 3.15 Solution structure of apo *v*-Src SH2. *Top left* | Backbone trace of ensemble, superposed according to N, C_α, C′ and O atoms (rms deviation from mean = 0.78 ± 0.09 Å). *Top right* | Cartoon representation of lowest-energy conformer. The secondary structure was defined by *MOLMOL*. *Bottom left* | Cartoon representation of mean structure, with tube thickness determined by the per-residue ensemble-averaged rms deviation from the mean backbone atom coordinates. Red, blue and grey colours refer to secondary structural elements – helices, strands and loops, respectively. *Bottom right* | Ramachandran plot of backbone angles for the ensemble; glycines are represented by crosses, all other residues as dots. This figure was prepared using *MOLMOL*[255].

All NOE violations in excess of 0.2 Å correspond to long range restraints between either amides or aliphatic protons and methyl groups, specifically the side chains of hydrophobic residues L20, V45, L54, I61, L64 and L98. These residues do not cluster at a single position within the structure, nor are they exclusively solvent-exposed or buried. No significant violations are found in residues that make up the binding site for phosphopeptide.

Restraint type (number)	**Mean**	±	**SD**
Unambiguous NOE distance restraints (2301)			
rmsd (Å)	0.019	±	0.001
no. violations > 0.5 Å	0	±	0
no. violations > 0.3 Å	1.4	±	1.2
no. violations > 0.1 Å	59.0	±	12.2
Ambiguous NOE distance restraints (454)			
rmsd (Å)	0.035	±	0.013
no. violations > 0.5 Å	0.3	±	0.4
no. violations > 0.3 Å	0.9	±	1.0
no. violations > 0.1 Å	8.9	±	2.0
All NOE distance restraints (2755)			
rmsd (Å)	0.020	±	0.002
no. violations > 0.5 Å	0.3	±	0.4
no. violations > 0.3 Å	2.3	±	1.4
no. violations > 0.1 Å	67.9	±	12.7
Hydrogen bond distance restraints (38 pairs)			
rmsd (Å)	0.013	±	0.004
no. violations > 0.5 Å	0	±	0
no. violations > 0.3 Å	0	±	0
no. violations > 0.1 Å	0.3	±	0.5
All distance restraints (2793)			
rmsd (Å)	0.020	±	0.002
no. violations > 0.5 Å	0.3	±	0.4
no. violations > 0.3 Å	2.3	±	1.4
no. violations > 0.1 Å	68.2	±	12.8
J-Coupling restraints (95)			
rmsd (Hz)	1.11	±	0.06
***TALOS* angle restraints** (21)			
rmsd (°)	12.40	±	1.75
no. violations > 5°	3.7	±	0.5

Table 3.11 Summary of restraint statistics

The stereochemical quality of backbone angles is assessed in the Ramachandran plot (Figure 3.15). *PROCHECK-NMR* analysis showed that over 83% of ϕ/ψ angle pairs were in the most favoured region of the graph, as defined by the program. A further 14 % fell within the additionally allowed region. The variability in backbone N, C_α, C′ and O atom positions across the ensemble was examined by calculating the rms deviation from mean Cartesian coordinates; this is shown graphically in Figure 3.15. Whilst secondary structural regions were associated with low backbone rmsd values (0.25 ± 0.04 Å), the global ensemble average was 3-fold higher (0.78 ± 0.09 Å). Deviation in the position of heavy atoms from mean coordinates

is compared with restraint density in Figure 3.16. On average, residues were involved in 35.3 ± 15.5 unique and unambiguous restraints*, although in regions of secondary structure an increased density was observed (42.8 ± 13.8). Loop and terminal regions displayed a below-average number of restraints per residue.

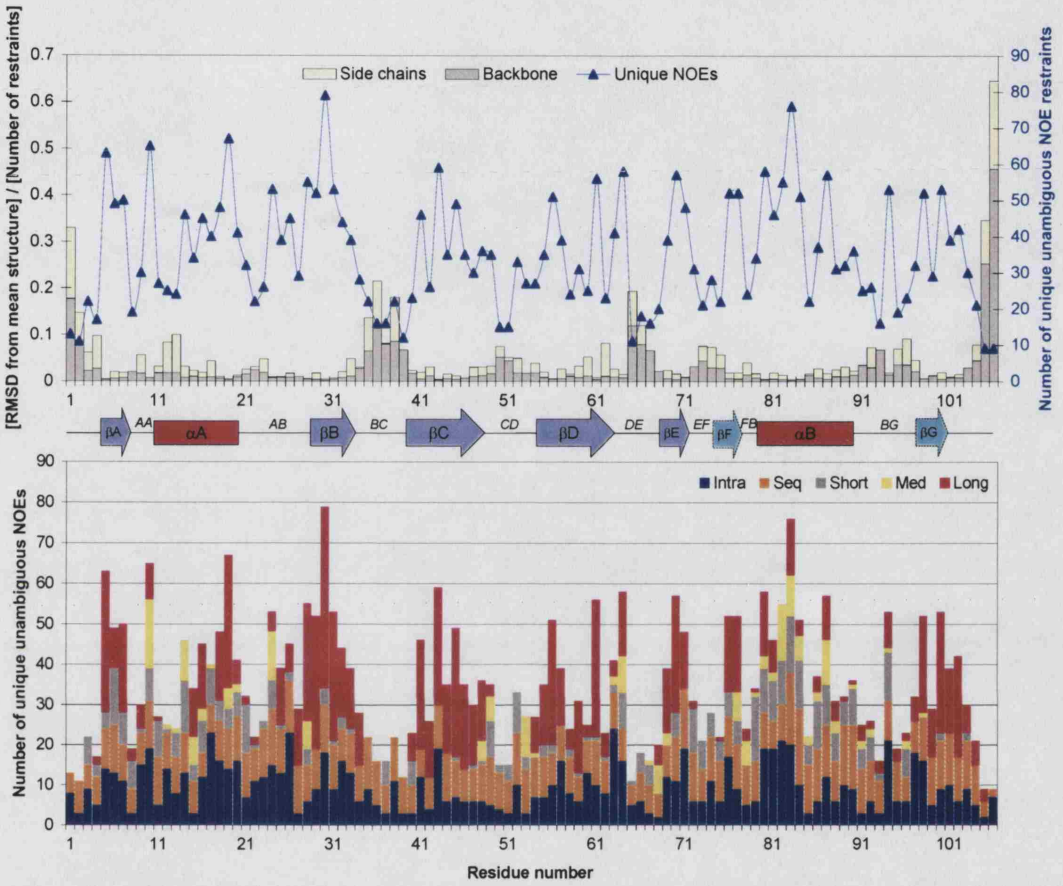


Figure 3.16 Per-residue ensemble statistics and secondary structure. *Top* | Ratio of the ensemble-averaged rms deviation for backbone (grey bars) and side chain (yellow bars) to the number of unique unambiguous NOE restraints. The total number of NOEs is shown as a blue line. *Bottom* | Breakdown of per-residue restraints according to distance in sequence (intra-residue, sequential, short-range (2-3), medium range (4-5) and long range (> 5)).

By classifying restraints according to their distance in sequence – intra-residue, sequential, short-range (2-3), medium range (4-5) and long range (> 5) – it is clear that regions showing relatively large coordinate rms deviations have fewer medium and long range restraints than well-defined parts of the domain (Figure 3.16). For instance, the BC loop, which forms part of the phosphotyrosine binding site, shows

*The final restraint list contained 2310 unambiguous and 454 ambiguous NOEs. The analysis software *AQUA* and *PROCHECK-NMR* cannot deal with ADRs. Moreover, any redundant restraints involving identical atomic pairs are removed, leaving a single unique restraint. Thus, the number of restraints per residue quoted here includes only unique and unambiguous contacts.

an above-average rms deviation because it is defined by relatively few restraints, none of which are within the medium or long range classes. Plotting the total number of restraints against the heavy atom rms deviation for each residue reveals a weak negative correlation (Figure 3.17). The graph shows residues defined by more than 40 restraints are generally associated with low backbone rms deviations (0.25-0.50 Å). The provision of more restraints appears to have minimal effect on this value. For residues with less than 30 restraints (and particularly less than 20), backbone rms deviations tend to increase. For side chain heavy atoms, it appears that 60-80 restraints per residue are required to achieve a high degree of precision (rmsd ~0.5 Å).

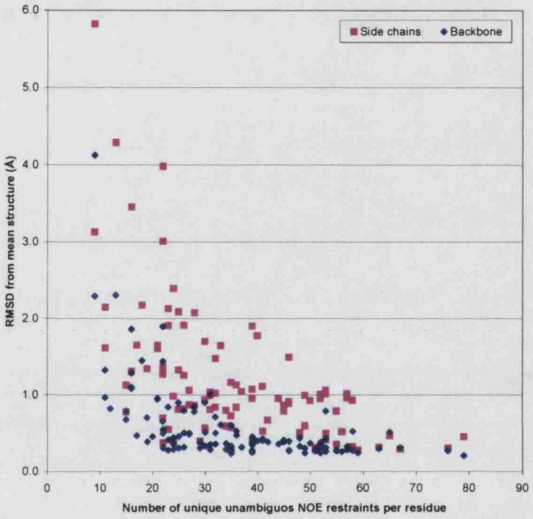


Figure 3.17 Scattergraph comparing the number of restraints per residue with backbone (blue) and side chain (magenta) rms deviations from mean heavy atom coordinates.

The correlation between restraint density and precision is not readily interpretable due to the large amounts of scatter, particularly in the side chain data set. It should be noted that other restraints not included here – ambiguous distances, hydrogen bonds, and angles – also contribute to the observed rms deviations. However, a statistical analysis of these data indicated that 85 % of residues with 30 or more restraints display backbone rms deviations below 0.5 Å and 70 % show side chain rmsds below 1.0 Å. This could provide a rough target for manual NOE assignment projects.

The circular variance of backbone ϕ/ψ angles provides another measure of precision by analysing the spread of these angles across the ensemble (Figure 3.18). Again, loops and termini are identified as regions in which significant difference occurs between structures.

The completeness of NOE data (Equation 3.19) is a more robust measure of data

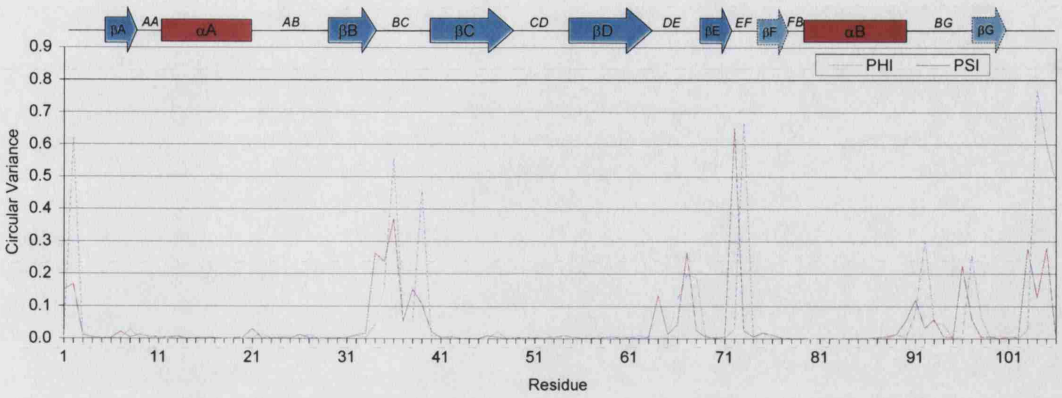


Figure 3.18 Plot of the circular variance of ϕ/ψ angles.

quality than the number of NOEs per residue, as it is normalised by the number of expected NOEs[261]. A list of expected NOEs is created by identifying interproton distances within the ensemble that are less than a user-specified cut-off. The number of matches between this list and the observed NOE list is then counted, and divided by the total number of expected NOEs. The ratio is commonly presented as a percentage of completeness.

$$\text{Completeness} = \frac{\text{number of matched observed NOEs}}{\text{number of expected NOEs}} \tag{3.19}$$

The program *AQUA*[254] was used to analyse the completeness of *v*-Src SH2 NOE data against the ensemble of structures. Using default parameters, the completeness of inter-proton distances up to 4 Å was 56 %. For comparison, in a group of 97 NMR-derived protein structures the average completeness value was 48 ± 13 %, whilst one structure achieved 75 %[261]. Completeness of data for *v*-Src SH2 is therefore above average, but within the normal range observed. It should be noted that ambiguous assignments and unambiguous assignments consisting of two or more contributions are not included in this analysis. Thus some expected NOEs are apparently not observed in the data. If all the data were taken into account, NOE completeness would be higher than 56 %. Statistical data regarding the completeness of NOE data for *v*-Src SH2 are shown in Figures 3.19 and 3.20.

In conclusion, the solution structure of *v*-Src SH2 has been determined to a high degree of precision and quality. Regions that display relatively high rms deviations from the mean structure are localised to loop and terminal sequences. Analysis of the input data revealed a below-average restraint density and a lack of medium to long range restraints within these regions. The completeness of NOE data was not systematically lower for loops and termini, suggesting that insufficient assignment was not the source of this variability. Instead it is highly probable that the relatively high rms deviations for these regions reflect conformational mobility.

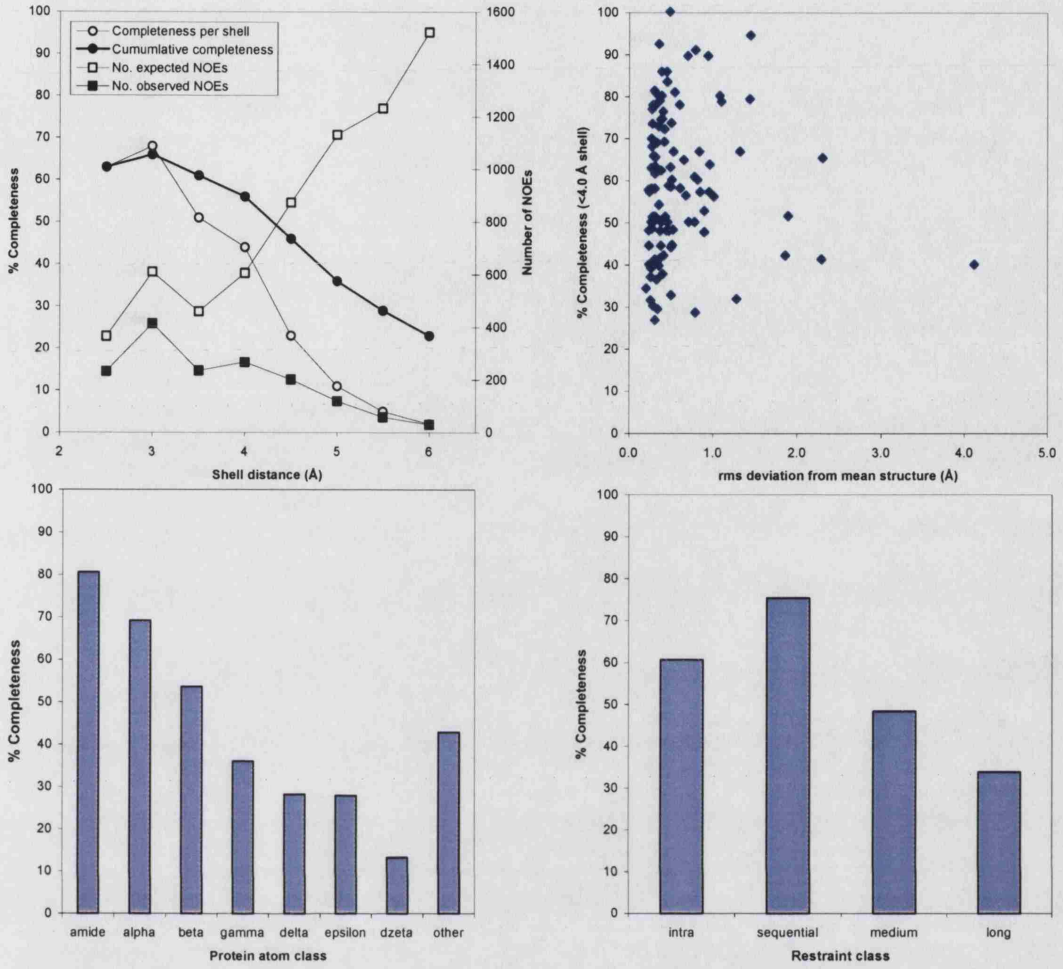


Figure 3.19 NOE completeness data for *v*-Src SH2. *Top left* | The % completeness and number of NOEs are compared for different cut-off values (2.5–6.0 Å). Completeness percentages are also shown for several different proton types (*bottom left*) and restraint classes (*bottom right*). *Top right* | Per-residue completeness data is compared to rms deviations from the mean ensemble structure.

3.7.1 Comparison of solution and crystal structures

In this section, the ensemble of solution structures is compared to MolA of the x-ray crystal structure (PDB code 1SPR) using structural alignments and contact maps. MolA was chosen because it displays the fewest contacts within the asymmetric unit of the crystal. The variation in atomic coordinates between the NMR ensemble and crystal structure was analysed in several ways (Table 3.7.1). Backbone heavy atoms in the range 2-104 were used to perform pair-wise alignments between individual conformers and the crystal structure. The mean global heavy atom deviation for backbone atoms was 2.01 ± 0.13 Å, which indicates a strong degree of global homology between the structures. Side chain positions showed more extensive deviations (3.38 ± 0.15 Å).

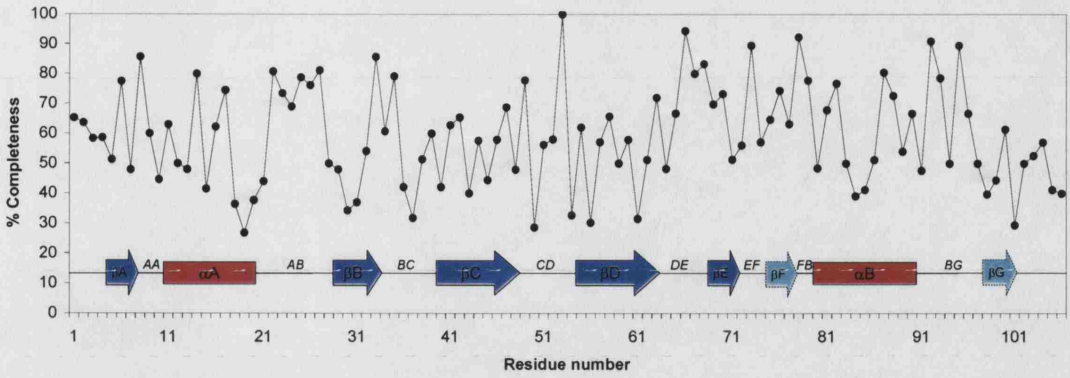


Figure 3.20 Per-residue NOE completeness with secondary structure shown inset for comparison.

Mean pairwise global rms (Å)	
heavy, backbone	2.01 ± 0.13
heavy, side chain	3.38 ± 0.15
all heavy	2.84 ± 0.13
Mean local rms (Å)	
heavy, backbone	0.39 ± 0.21
Mean local displacement (Å)	
heavy, backbone	0.34 ± 0.20
heavy, side chain	1.80 ± 0.99
all heavy	1.35 ± 0.75

Table 3.12 Comparative statistics describing the NMR and crystal structures of *v*-Src SH2. For a description of 'global' and 'local' alignments, see the main body of text. To obtain local rms and displacement scores, residue triplets were superimposed, and the deviation calculated for the triplet and the central residue, respectively. Statistics were generated using *MOLMOL*.[255]

Global alignment algorithms attempt to find the best fit between two structures according to a set of atoms across all residues. Local alignment methods perform a similar operation, but with much shorter segments of the protein, typically 3-5 residues. This allows one to distinguish between large per-residue rms deviations caused by significant conformational differences and those caused by systematic displacement of essentially equivalent conformations. The mean local rms deviation, obtained by fitting residue triplets and averaging the rms score across the three residues, for backbone atoms was 0.39 ± 0.21 Å. This suggests that when considered in isolation, short sequences of the domain are very similar. The side chain deviations (and range of deviation) are significantly larger, indicating differences between the structures derived by the two methods. The alignment of NMR and crystal structures and the magnitude of local rms differences are shown in Figure 3.21. Secondary structural regions display relatively small local rms deviations, whilst loops and termini show greater variability.

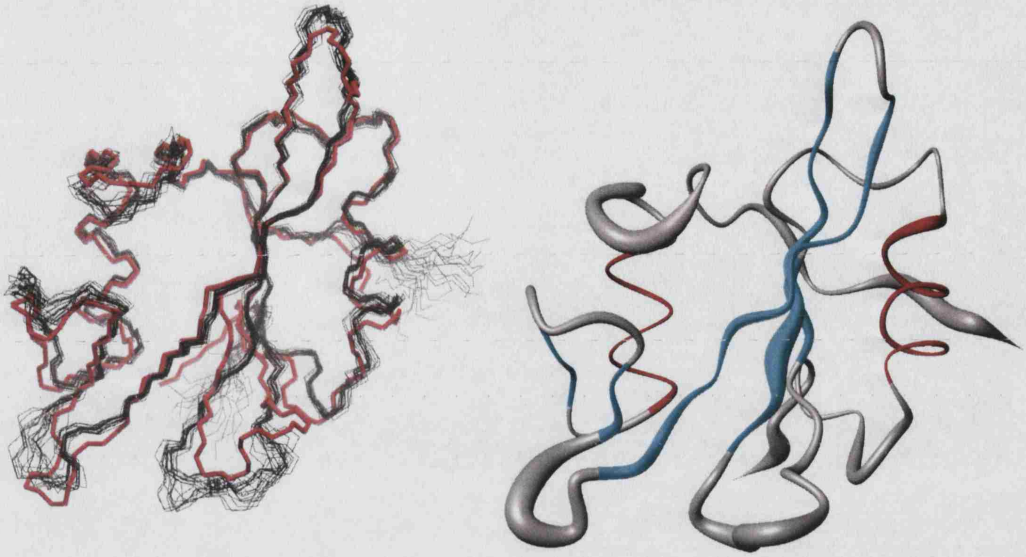


Figure 3.21 Superimposition of the NMR ensemble on the X-ray crystal structure. *Left* | Global alignment between individual ensemble conformers and the x-ray structure using backbone heavy atoms. *Right* | Tube representation of lowest-energy ensemble structure with spline diameter related to the local rms deviations in backbone atoms of ensemble structures from the crystal structure. α-helical and β-strand regions are highlighted with red and blue colours, respectively. This figure was prepared using *VMD*[262] and *MOLMOL*[255].

The NMR and crystal structures were also compared using contact maps (Figure 3.22), which are useful descriptors of inter-residue proximity. Both the structure alignment and contact maps indicate that the structures differ at positions 41-44 (βC strand), which forms part of the pY binding site. The distance between residues 9-11 and 41-44 is about 2-3 Å larger in the crystal structure and there is a visible 'kink' in the βC strand. It is quite possible that binding of phosphate to the crystal structure could lead to this change. If this structure was also formed in the apo solution state then NOEs would be expected to occur between the side chain methyl groups of L43 and L94 and I71. These NOEs are not observed in the apo ^{13}C-NOESY spectrum. It would be interesting to discover if they are present in the an equivalent phosphate-bound spectrum.

The contact maps also show a decreased separation between the DE and BG loops (corresponding to residues 71-73 and 92-95, respectively) in the solution structure compared to the crystal structure. This appears to reduce the size of the pY+3 pocket. It is not clear what the reason for this difference is. There is an inter-domain hydrogen bond between MolA/D92 and MolC/R13 in the crystal structure and it is possible that interactions between the two domains could lead to a greater distance between the loops. Additionally, the displacement of residues 41-44 towards the pY+3 pocket may have a knock-on effect on DE and/or BG loop residues.

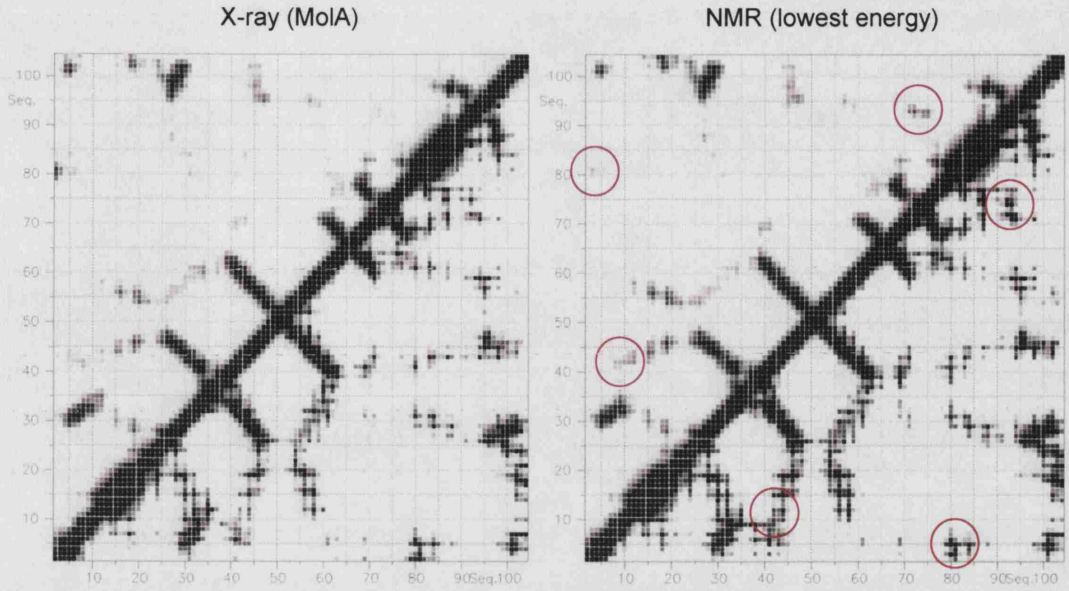


Figure 3.22 Contact maps describing the X-ray (*left*) and NMR (*right*) structures of *v*-Src SH2. Pairwise heavy atom contacts are plotted for residues 2-104. Contacts between backbone atoms shown above the diagonal whilst contacts between all heavy atoms are illustrated below the diagonal. Distances below 4 Å are coloured black and a gradient of grey-white is used to describe distances in the range 4-10 Å. Red circles highlight some of the slight differences between the two maps. This figure was prepared using *MOLMOL*[255].

In conclusion, it has been shown that the backbone structure of *v*-Src SH2 is very similar in the crystal and solution states; the only major difference (C42/L43) is probably due to the presence of phosphate rather than crystallographic conditions *per se*. These findings support the assertion that the NMR structure presented here is a more accurate representation of the apo state than the peptide-free crystal structure.

3.8 Analysis of automated NOESY assignment

Clearly, *ARIA* is capable of generating a good quality ensemble of structures and it is the purpose of this section to analyse the path of *ARIA* in reaching this state. The final *ARIA* run, consisting of 9 rounds of iterative assignment, utilised NOEs from four NOESY spectra, a list of hydrogen bonds, and a list of unambiguous restraints from the previous *ARIA* run, as well as angular restraints in the form of *TALOS*-derived dihedrals and $^3J_{HNH\alpha}$ couplings. During the 9 iterations, changes in assignment parameters lead to a gradual reduction in NOE assignment ambiguity (Table 3.13). In the final iteration, nearly 84 % of assignments are unambiguously defined. Since identical assignments are merged, an apparent decrease in the total

number of NOEs is observed. Unambiguous NOE restraints were also classified according to the relative sequence positions that they connect. The final list consists of 1014 intra-residue, 511 sequential, 216 short range (2-3), 90 medium range (4-5), and 470 long range (> 5) restraints.

	^{15}N (600 MHz)		^{15}N (800MHz)		^{13}C (aliphatic)		^{13}C (aromatic)		Total	
Iteration	N_{un}	N_{am}	N_{un}	N_{am}	N_{un}	N_{am}	N_{un}	N_{am}	N_{un}	N_{am}
0	4	192	86	1561	131	2180	1	25	222	3958
1	69	302	147	1213	361	1893	2	16	579	3424
2	148	197	308	959	681	1389	7	10	1144	2555
3	215	104	585	615	949	836	8	6	1757	1561
4	230	86	655	522	1020	680	7	6	1912	1294
5	240	70	732	416	1080	540	8	5	2060	1031
6	253	54	795	314	1084	434	8	5	2140	807
7	249	46	819	274	1113	363	7	5	2188	688
8	259	32	869	181	1164	239	9	2	2301	454

Table 3.13 Iteration assignment statistics for each NOESY spectrum. The 600 and 800 MHz ^{15}N-NOESY data correspond to the short and long mixing time spectra, respectively, and N_{un} and N_{am} refer to the number of unambiguous and ambiguous assignments, respectively. As the number of unambiguous assignments increases, duplicates within the data also increases, in a quadratic fashion. These redundant restraints are removed, hence the total number of assignments appears to decrease over the 9 iterations.

The change in assignment statistics is also shown graphically in Figure 3.23. The most dramatic changes occur between Iterations 0-3. This is probably because the inclusion of a list of trusted unambiguous restraints leads to well-defined structures early on in the iterative process. Thus ambiguities in the data are resolved earlier than if these supplementary restraints were excluded. The trends observed in Figure 3.23 do not plateau by iteration 8, suggesting that the standard *ARIA* protocols could be modified to include further iterations which may reduce ambiguity further.

The ambiguity histrograms shown in Figure 3.23 indicate that the number of possible contributions to each assignment are reduced steadily over the 9 iterations. In iteration 8, approximately 90 % of ambiguous assignments consist only two possibilities, while 9 % involve three contributions. These results show that *ARIA* is able to resolve highly ambiguous assignments through iterative structure calculations and re-assignment protocols.

In order to assess the change in structural precision and quality over the 9 iterations, a small ensemble of conformers was chosen from within the larger set of results for each iteration. Structures were ranked according to the simulated annealing NOE energy term (E_{noe}), which reflects the total amount of NOE violation occurring. The five lowest-energy structures were selected for analysis. The rms deviation from NOE data is reported in each structure file and was simply averaged across the five structures. The deviation between the backbone atomic positions

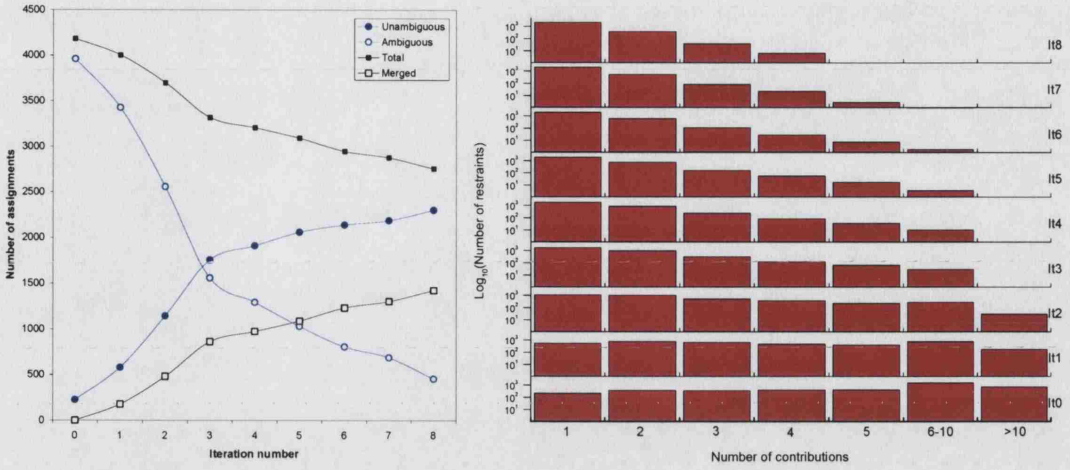


Figure 3.23 Changes in the assignment ambiguity with iteration number. *Left* | The population of (un)ambiguous assignment classes are inversely related, while the total number of restraints decreases due to merging of equivalent restraints. *Right* | Ambiguity histograms show a gradual decrease in the number of possible contributions to each assignment.

in these five structures and equivalent atoms in the final mean structure were calculated using *PyMOL* (*DeLano Scientific*). Structural quality was also assessed by analysing the distribution of ϕ and ψ backbone angles (*PROCHECK-NMR*[254]). Figure 3.24 displays the progressive convergence of iterative ensembles towards the final ensemble in terms of the rms deviations from the final ensemble mean coordinates and from the NOE data. Due to the provision of unambiguous restraints from the previous *ARIA* run, the best five structures in the first iteration already show a low backbone rms deviation from the final mean structure.

The deviation from NOE distance restraints shows continuous decrease over the 9 iterations, although a slight increase is found in the water-refined structure ensemble. Interestingly, this trend is correlated in a linear fashion to $\log_{10}$ of v_{tol}, the violation tolerance parameter, suggesting that inconsistent restraints are rapidly identified and dealt with, leading to smaller deviations from the input data.

The quality of structures, as measured by Ramachandran angle analysis, also increases with iteration number (Figure 3.24). The proportion of angles that are localised to the 'most favoured' regions starts at fairly high levels (~70 %). It is highly likely that addition of supplementary data – unambiguous distances, *TALOS* angles, and J-couplings – is the reason for these high quality structures, despite the fact that assignments are highly ambiguous at this early stage. Unlike the rms deviations from NOE data and final coordinates, structural quality does not appear to be strongly influenced by *ARIA* assignment parameters, such as p or v_{tol}.

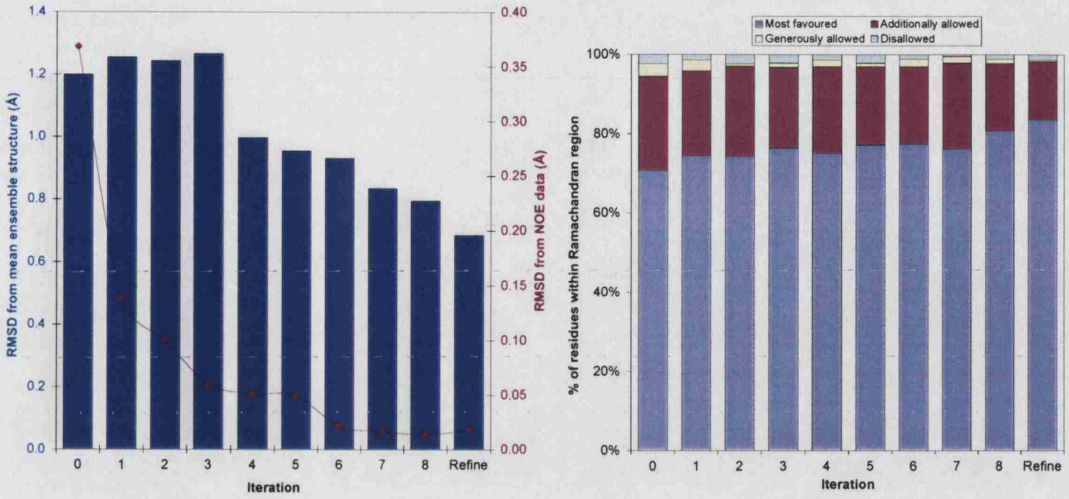


Figure 3.24 Changes in structural precision and quality over the 9 iterations. *Left* | Rms deviations of backbone atoms from the final ensemble mean coordinates (blue bars) and from the NOE data (brown line). *Right* | *PROCHECK-NMR* Ramachandran analysis of iteration ensembles. Backbone angles were classified according to their location in the Ramachandran plot.

3.8.1 Evaluation of *ARIA*

This structure determination was the first time our group had employed automated NOESY assignment methods. The exploratory use of *ARIA* has highlighted several areas which demand careful attention. Firstly, the success of iterative assignment and structure calculations is strongly dependent on the quality of the input data – chemical shifts, NOEs, and supplementary restraint information. In order to achieve correct assignments, all spectra should be referenced consistently. The definition of chemical shift tolerance windows is also important. Large windows allow poorly aligned spectra or badly-matched NOEs to be correctly assigned, however the overall level of ambiguity remains high and convergence is negatively affected. Conversely, narrow chemical shift tolerances lead to missed assignments, resulting in an incorrect structure and loss of restraints. Recently, the authors of *ARIA* published a method for optimisation of chemical shift tolerances and the N_{max} parameter[253]. The new script also facilitates the identification of NOEs that match badly with the suggested assignment, which aids in the quality control of NOESY peaks. Such developments are welcome as they seek to improve the quality of input data and the efficiency of its use.

The NOESY data describing the *v*-Src SH2 domain was assigned partially by hand prior to application of *ARIA*. This enabled validation of NOE peak positions, especially in the case of ^{15}N spectra, which was assigned to a greater extent than ^{13}C data. As a result, in the majority of cases where an unlikely restraint was vio-

lated severely, the error corresponded to spectral regions in which previous manual assignment was sparse. Therefore manual verification of crosspeak locations is a useful method of generating high quality NOESY data. Unfortunately, this is a slow process that could negate attempts to use *ARIA* for large-scale protein structure determination. It also detracts from the objectivity of automatic assignment methods. Improvements in automated peak placement software will help to resolve this issue, but currently there does not seem to be a reasonable alternative to manual validation of NOESY data before, if not during, the application of *ARIA*. The program aims to discard erroneous assignments that are inconsistent with the bulk of data, and with each other, however there have been no reports concerning what proportion of the data needs to be correct in order for this to occur. For this *v*-Src SH2 structure determination, it is estimated that 15-20 % of NOE peaks in the initial list had to be adjusted or discarded in order to achieve convergent structures of high quality. In conclusion, although *ARIA* facilitates automated assignment of NOESY spectra, the accuracy of such assignment is highly sensitive to iteration parameters and the quality of input data. As significant human intervention is required, the elucidation of reliable NMR structures using this approach remains, at best, semi-automatic.

The quality and provision of supplementary data is also a matter for concern. *TALOS*-derived phi/psi angle predictions are used routinely in NMR structure determination projects. Whilst the relationship between backbone chemical shift patterns and angular conformation is reasonably well characterised, the error bounds for restraint satisfaction had to be set quite large ($\phi/\psi \pm 50°$) in this project, suggesting that angular predictions contained many errors. In this regard, *TALOS* confidently predicts values for only 65 % of a given protein, however it is claimed that 97 % of such estimates turn out to be correct[70]. In the case of *v*-Src SH2, 58 % of residue predictions were classed as 'Good', but only 20 % of these proved to be consistent with the NOE data and were included in the final structure calculations. There is a temptation to use *TALOS* predictions in early *ARIA* runs in order to improve convergence and facilitate the identification of inconsistent NOEs. However, this is dangerous as provision of incorrect torsion angles will be inconsistent with both good and bad NOE data.

A similar situation exists for hydrogen bond distance restraints, which are predicted on the basis of small amide proton solvent-exchange rates and temperature coefficients, or consistent structural information. Again, it is advisable use *ARIA* to obtain good quality NOESY data prior to the introduction of these restraints.

Although the central function of *ARIA* is assignment of ambiguous NOEs, it would be useful to modify the program to include intelligent use of unreliable supplementary data. For instance, the first few iterations could rely on primarily NOE

data, although other empirical data (i.e. J-couplings, residual dipolar couplings, and chemical shifts) could be included at this stage. Once a certain threshold of convergence has been reached within iterative ensemble calculations, restraints derived from predictions (dihedrals and hydrogen bonds) could then be validated against the structures, and used appropriately in subsequent iterations. In the existing version of *ARIA*, the first iteration in which supplementary data are introduced is a user-defined parameter, but there are no checks as to whether the data are broadly consistent with NOE-derived ensembles.

ARIA v2.0 uses floating assignment protocols to deal with prochiral protons and the methyl groups of leucine and valine. Whilst this was appropriate for this SH2 domain, other projects may include stereospecific assignments. It is possible to specify such assignments, but in a non-user-friendly manner. It was also discovered that *ARIA* applies floating assignment to side-chain amine protons and the methyl groups of isoleucine. This complicates the NOE assignment problem unnecessarily. XML code describing these protons was altered in the chemical shift definition file, however this has to be repeated whenever the chemical shift lists are changed, or a new *ARIA* run is prepared. Future versions of *ARIA* will provide a user-friendly system for specifying the extent of stereospecific assignment (Michael Habeck, program co-author, personal communication).

3.9 Hydrogen bonding patterns within *v*-Src SH2

NMR has been used widely in the identification of hydrogen bonds within molecular structures such as small polypeptides[263], globular proteins[264], and nucleic acids[265]. Although hydrogen bonds are crucial to understanding biomolecular interactions, their characterisation has relied until recently upon the spatial proximity of acceptors and donors[241] coupled with indirect supporting evidence in the form of amide proton solvent-exchange rates[266] and temperature coefficients[267]. In the last few years, NMR pulse sequences have been optimised in order to measure the small J-coupling across the hydrogen bond[242]. This not only enables unambiguous identification of the hydrogen bond donor and acceptor, but also provides quantitative data regarding bonding geometry. At current sensitivity levels, such experiments require stability of the sample at high concentrations (> 1 mM) over several days. These conditions precluded direct observation of hydrogen bonds within the *v*-Src SH2 domain. The indirect approaches to investigating hydrogen bonding – resistance to solvent exchange and low temperature sensitivity – have lower sample and temporal requirements and have been used extensively, particularly the former, in protein structure determination projects.

3.9.1 H/D exchange rates

The theoretical and practical aspects of measuring the exchange rates of amide protons has been well-characterised[266]. Typically, the exchange between protonated amides and deuterated water is followed over time and fitted to an exponential curve. Amides that are buried, or involved in hydrogen bonds generally display lower solvent exchange rates than their surface-exposed counterparts. There are several weaknesses to this approach that complicate the relationship between exchange rates and hydrogen bonds. Firstly, Pedersen *et al.* demonstrated that the distance between a given amide and the solvent-accessible surface was a strong determinant of exchange rate[268]. Thus, amides forming intramolecular hydrogen bonds that are at or near the surface and buried, non-hydrogen bonded amides can display high and low solvent exchange rates, respectively. Solvent-exchange rates were also shown to be dependent on pH[268] and fluctuations in local structure[269]. As this approach relies on aqueous solvent exchange, it is necessary that the molecule in question exhibit relatively high solubility. For certain polypeptides that are more readily soluble in aprotic solvents, this issue impedes analysis[270].

3.9.2 Amide proton temperature coefficients

An alternative approach to the identification of hydrogen bonds involves interpretation of the temperature dependence of amide proton chemical shifts. Protein structure and dynamics are sensitive to surrounding thermal energy[271, 272]. Tilton *et al.* demonstrated that the Ribonuclease-A structure expands linearly with increasing temperature (0.4 % per 100 K) due to changes in atomic packing and bond lengths[273]. The number of sub-3.25 Å inter-atomic distances within the crystal structure was observed to decrease by 2.7 % between data sets collected at 98 and 300 K. When this value was separated according to atom type, the number of sub-3.25 Å distances involving oxygen or nitrogen decreased by 2.4 %, those involving non-polar carbon and sulphur atoms decreased by 3.1 %, and finally, the number of potential intramolecular hydrogen bonds decreased slightly by 1 %.

Given the fact that proteins expand slightly with temperature, it is expected that the NMR spectrum would change as temperature is altered. Indeed, Kopple *et al.* demonstrated that amide protons within small peptides show temperature-dependent chemical shifts, with differential (but still linear) sensitivities observed for intramolecular (internal) hydrogen-bonded amides compared with their solvent-exposed (external) counterparts[274]. Baxter & Williamson noted that temperature-dependent changes in globular protein chemical shifts were generally linear up to 15 °C below the global denaturation temperature[270].

The chemical shift of an hydrogen-bonded amide proton is strongly dependent on the proximity of the electronegative acceptor atom[†275]. As the distance between the proton and acceptor increases with temperature, the proton nucleus becomes more shielded against the local electromagnetic environment resulting in a negative change in chemical shift. This migration in shift with temperature can be fitted to a linear equation in which the gradient represents a temperature coefficient, in units of parts-per-billion per Kelvin (ppb/K). Temperature coefficients have been used to predict hydrogen bond donors, with values more positive than -5 ppb/K suggesting involvement in an intramolecular hydrogen bond[270].

Cierpicki & Otlewski performed a recent statistical analysis of amide temperature coefficient data for 14 proteins, each with high-resolution solution structures[267]. The average temperature coefficient displayed by amides forming intramolecular hydrogen bonds was -3.2 ± 2.0 ppb/K whereas for amides with no obvious hydrogen bond acceptor, this decreased to -7.1 ± 2.5 ppb/K. Over 85 % of amides with temperature coefficients above -4.6 ppb/K were involved in intramolecular hydrogen bonds. By considering the range -4.0 → -1.0 ppb/K, this proportion increased to 93 %. Temperature coefficients were also dependent on secondary structure – 70 % of amides in α-helices had values above -4.6 ppb/K; equivalent values for β-sheet, 3_{10}-helices, and β-turns were 87, 90, and 98 %, respectively.

Temperature coefficients are not found to be correlated with surface exposure or pH, which is an advantage over analysis of H/D exchange rates[270]. However, two caveats that complicate the analysis of temperature coefficients are noted by Cierpicki & Otlewski[267]. First, in the case of very strong hydrogen bonds (short proton–acceptor distances) due to the inverse cube dependence of amide proton chemical shift on acceptor distance, temperature-induced expansion leads to large negative temperature coefficients, similar to those associated with solvent-exposed amides. Secondly, close proximity of an amide and an aromatic group can cause the deshielding effect of hydrogen bond lengthening to be attenuated by ring-current effects[‡] resulting in small temperature coefficients. Almost 90 % of amides whose chemical shifts were strongly affected by a neighbouring aromatic ring had temperature coefficients more positive than -2 ppb/K[267]. A third criticism of the reliability of temperature coefficients in predicting hydrogen bonding rests upon the fact that backbone amides are exquisitely sensitive nuclei. Therefore, temperature-induced chemical shift changes may represent multiple contributions from global and local structure and dynamics.

[†] H_N δ is related to $1/r^3$ where r is the distance between the amide proton and the hydrogen bond acceptor.

[‡] Slight alteration in the orientation and dynamics of aromatic rings can have a profound influence on local chemical shift values.

Amide temperature coefficients and H/D exchange rates both provide reasonably good predictions of hydrogen bonding in globular proteins. Although each technique is prone to complications that lead to incorrect predictions, the results are complementary. For instance, surface-exposed, intramolecular hydrogen bonded amides overlooked by H/D exchange can be identified using temperature coefficients. Conversely, very strong hydrogen bonds, or hydrogen bonds close to aromatic groups that are not faithfully reported in temperature coefficient data can be validated using H/D exchange rates. Baxter & Williamson compared data from both approaches and proposed that amides with relatively slow exchange rates and temperature coefficients more positive than -4.5 ppb/K should be considered as hydrogen bonded[270].

3.9.3 Measurement of temperature coefficients for *v*-Src SH2

This approach was used to explore hydrogen bonding within *v*-Src SH2 using the refined solution structure presented in this chapter to interpret results. Nine $^{1}H^{15}N$-HSQC spectra were recorded across the temperature range 10–34 °C (283–307 K), at 3 °C intervals (see Section 2.3.2 for spectral parameters). No attempt was made to monitor or correct for pH changes with increasing temperature as amide temperature coefficients are generally pH-independent[267]. The resultant spectra were processed as described in Section 2.3.3. Figure 3.25 displays the overlaid HSQC spectra. It is immediately clear that individual amides respond in a variable manner to temperature increases – the majority show a clear decrease in proton shift, whilst for others the decrease is only very marginal. Four amides (N21, D49, N55, Y86) display a positive correlation between temperature and proton chemical shift.

In order to calculate individual temperature coefficients, proton chemical shifts were obtained for each amide at each temperature using ANSIG[226]. Data for each amide was fitted to a linear slope using a Fortran-based routine kindly provided by Dr Matthew Cliff. The calculated temperature coefficients ranged between -9.5 and +5.4 ppb/K, with an average value of -4.0 ± 2.7 ppb/K.[§] Figure 3.26 displays exemplary behaviour for several amide proton chemical shifts.

The calculated temperature coefficients for the *v*-Src SH2 sequence (excluding prolines and N-terminus) are shown in Figure 3.27, and their distribution across the data range in Figure 3.28. Raw temperature coefficient values can also be found in Appendix C. On first examination, it appears that the values are randomly distributed across the sequence, apart from the terminal residues which clearly have, as expected, a strong negative temperature coefficient.

Temperature coefficient data were used to colour the apo solution structure of *v*-Src SH2, highlighting key regions of interest (Figure 3.28). Generally, amides

[§]Errors refer to the standard deviation

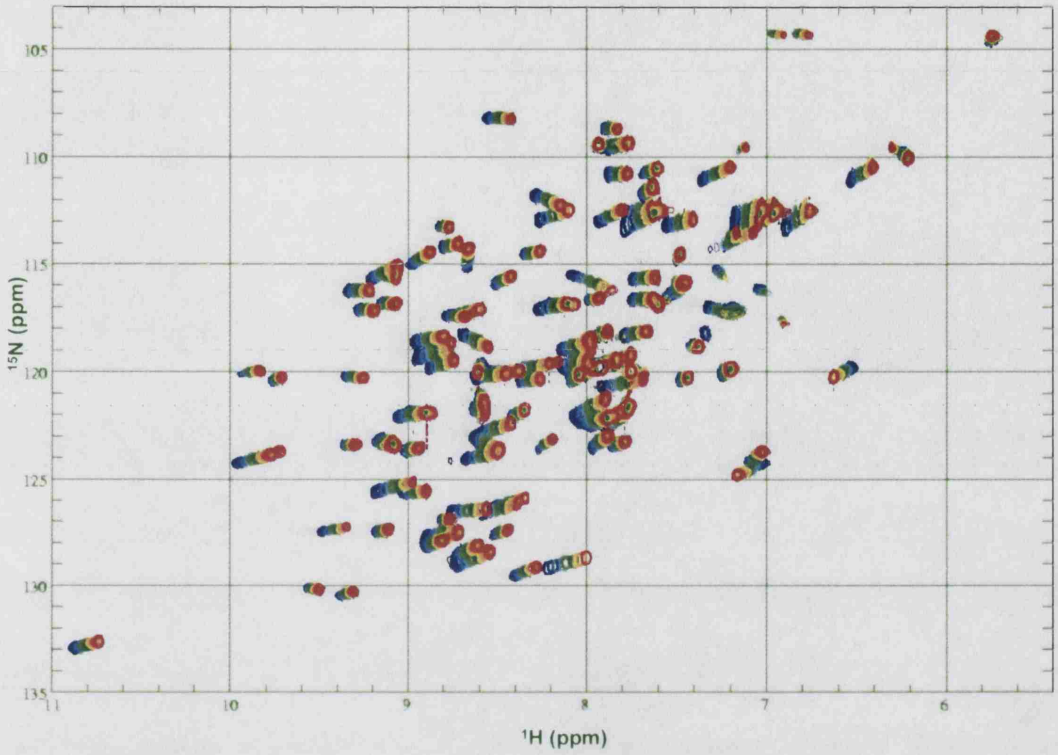


Figure 3.25 $^{1}H^{15}N$-HSQC spectra of *v*-Src SH2 recorded at 3°C intervals between 283 and 307 K. Spectra across this range are coloured according to a reversed visible light spectrum (blue → red).

buried in the domain core show small, negative temperature coefficients (white/pink colour), whilst values for solvent-exposed amides tend to be more negative (pink/red colour). There are some notable exceptions, however, which will be considered below. The alternating colour patterns along the front-facing, solvent-exposed βD-strand (running from the top right to bottom left of structure) reflects clearly the difference between exposed amides that hydrogen bond with bulk water (red), and those that form intramolecular hydrogen bonds within the β-sheet (white/pink). Using knowledge of the apo solution structure, temperature coefficient data were related to several characteristics of protein conformation – hydrogen bonding, secondary structure, proximity to aromatic groups and backbone dynamics.

3.9.3.1 Influence of hydrogen bonding patterns

As discussed above, hydrogen bonding patterns are a major determinant of temperature coefficients. The program *HBPLUS*[241] was used to identify hydrogen bonds in the *v*-Src SH2 solution structure ensemble. Temperature coefficient statistics relevant to amides forming internal and external hydrogen bonds can be found in Table 3.14. The average temperature coefficient for amide protons forming an in-

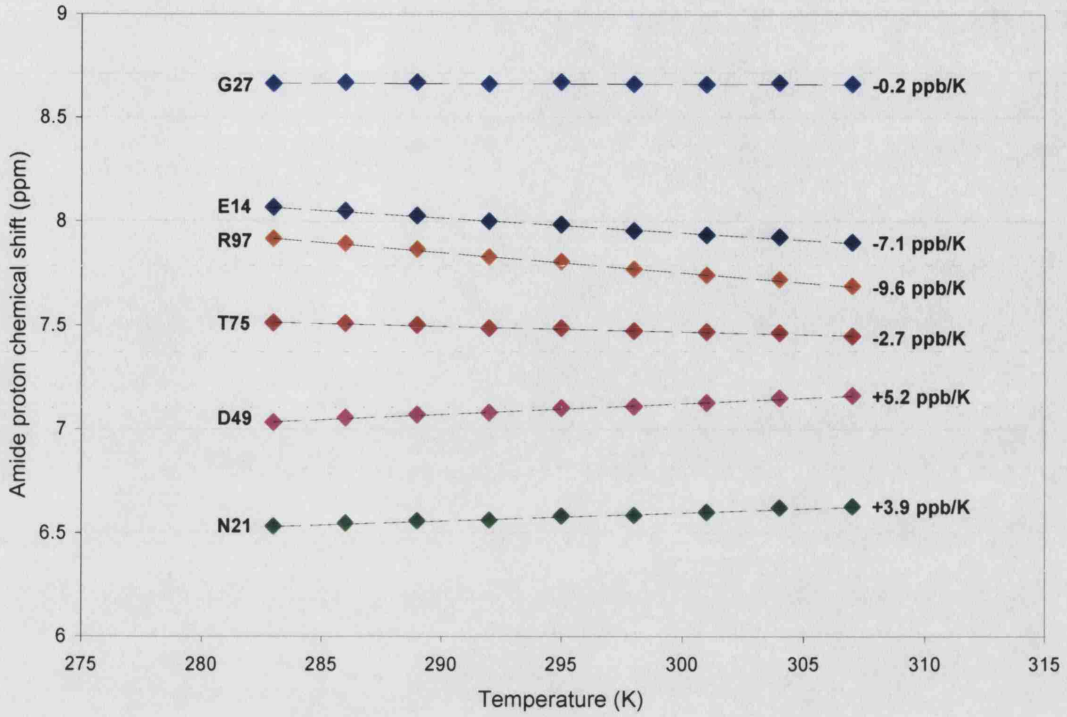


Figure 3.26 Temperature dependence of chemical shifts for selected amides. The linear slope of this change is the temperature coefficient.

		Secondary structure			Hydrogen bonding	
	Full data	α-helix	β-strand	Loop	Internal	External
Mean	-4.0	-3.6	-3.4	-4.6	-2.6	-5.2
Mean Error	0.27	0.46	0.32	0.50	0.30	0.34
St. Dev.	2.7	2.1	2.0	3.3	2.1	2.5
Min	-9.5	-7.0	-7.6	-9.5	-7.6	-9.5
Max	5.4	0.2	1.2	5.4	5.4	4.0
n	102	21	38	43	48	54

Table 3.14 Statistical analysis of temperature coefficient dataset according to secondary structure and hydrogen bonding patterns. Internal and external hydrogen bonding distinguishes between amides that form intramolecular bonds and those with no obvious partner, which are assumed to form hydrogen bonds with aqueous solvent. All values refer to temperature coefficients in units of ppb/K.

tramolecular hydrogen bond was -2.6 $\pm$ 2.1 ppb/K, whilst the remaining amides scored significantly lower at -5.2 $\pm$ 2.5 ppb/K. Compared with previously published statistics[267] (-3.2 $\pm$ 2.0 and -7.1 $\pm$ 2.5 ppb/K, respectively), *v*-Src SH2 is within the normal range of values, although the difference between the two populations was somewhat smaller the average.

Figure 3.29 displays the temperature coefficient data annotated according to the presence or absence of an intramolecular hydrogen bond. Data describing the former group are distributed in a bell-shaped curve around the average. Amides assumed to form hydrogen bonds with solvent are more dispersed, however the population is

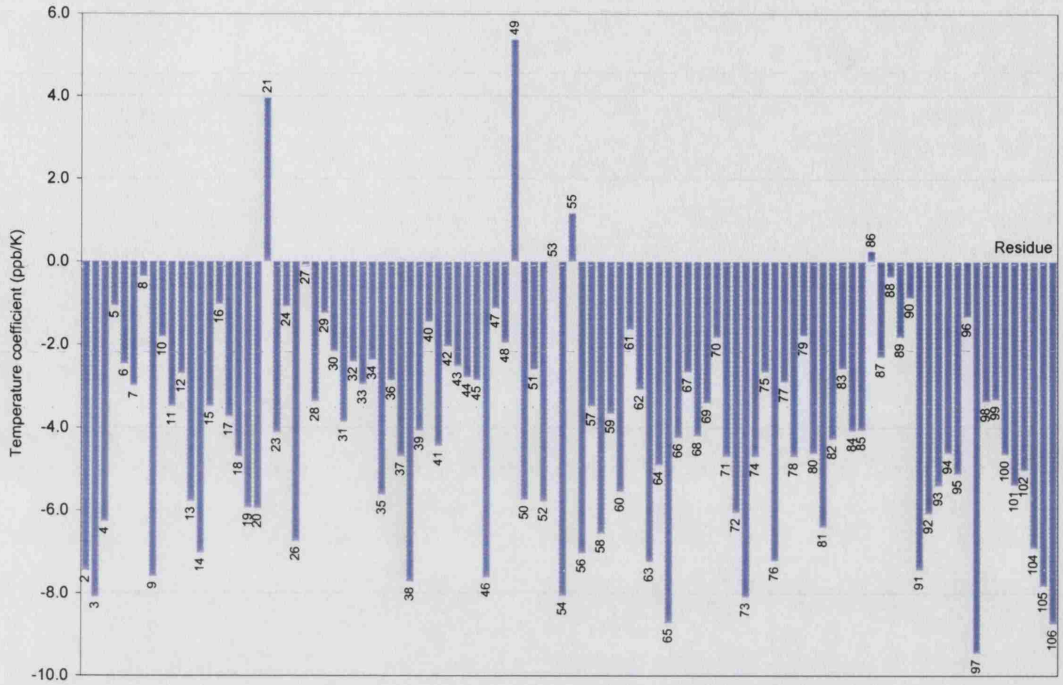


Figure 3.27 Temperature coefficients for *v*-Src SH2

clearly shifted to more negative temperature coefficients indicating that it correlates quite well with hydrogen bonding patterns. Cierpicki & Otlewski reported that, on average, 85 % of amides with temperature coefficients more positive than -4.6 ppb/K were involved in internal hydrogen bonding[267]. A similar analysis of *v*-Src SH2 data generated a value of 70 % (41/59). If the range -4.0 to -1.0 is considered (to ignore outliers caused by ring current effects), the published figure increases to 93 %, however that for *v*-Src SH2 remains significantly lower at just 71 % (30/42).

3.9.3.2 Influence of secondary structure

To explore the effect of secondary structure on temperature coefficients, amides were split into three categories according to whether they belonged to α-helices, β-strands, or loop/termini regions (Figure 3.29). It was found that the average temperature coefficient for amides in α-helices and β-sheets was slightly higher than the global average, whilst amides located in loop regions tended to show a broader range of more negative coefficients (see Table 3.14). Hydrogen bonding is clearly much stronger determinant of temperature coefficients than secondary structure. This is readily interpretable as the three amides at the N-terminus of an α-helix have no fixed hydrogen bonding partner, as do alternate amides in solvent-exposed β-strands (e.g. see residues 12-14 and 80-82). Cierpicki & Otlewski reported that 70 % of amides in α-helices displayed a temperature coefficient more positive than

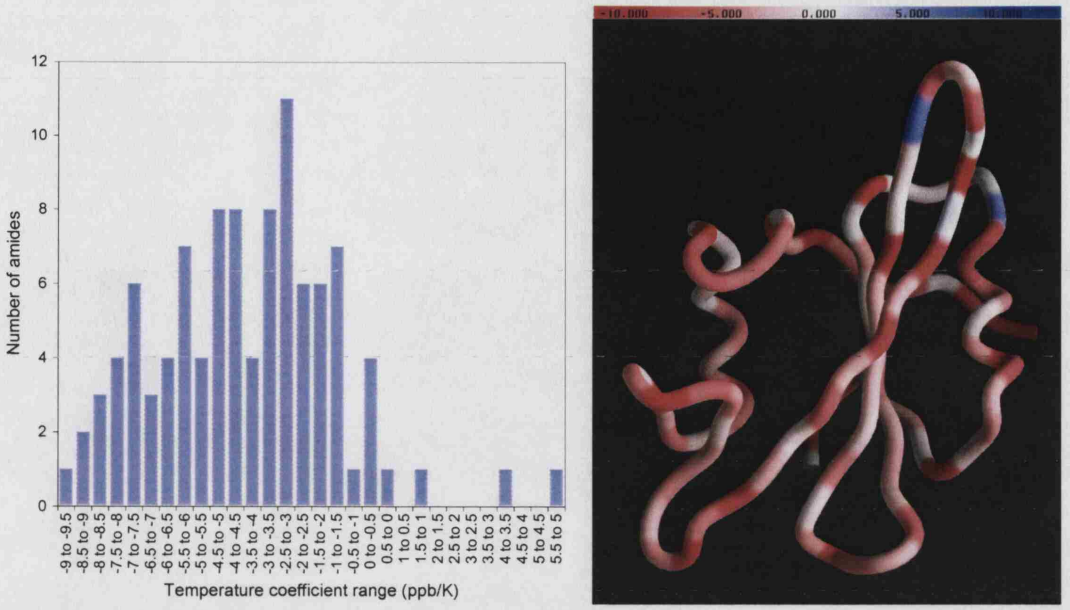


Figure 3.28 Temperature coefficient data analysis. *Left* | Histogram showing the distribution of temperature coefficients for *v*-Src SH2. *Right* | *v*-Src SH2 C_α backbone worm coloured according to temperature coefficient. This image was prepared using *GRASP*[276].

-4.6 ppb/K. The analogous figure for *v*-Src SH2 is 67 %, which agrees very well. For β-strands, the published figure was 87 %, however only 76 % β-strand amides exceeded the threshold in the *v*-Src SH2 data. This is possibly because in *v*-Src SH2, a significant proportion of amides in β-strands (~30 %) are likely to form hydrogen bonds with bulk water.

3.9.3.3 Influence of aromatic groups

Aromatic ring current effects are known to distort the chemical shift temperature dependence of neighbouring amides[267]. To investigate this phenomenon with the *v*-Src SH2 domain, backbone amides within 4 Å of an aromatic ring carbon were identified and their temperature coefficients analysed statistically (Table 3.15). It was found that proximity to aromatic rings does have a slight effect on the mean temperature coefficient, although there are clear examples of amides close to such groups that display large negative coefficients (e.g. Q81) and conversely, amides far from aromatic rings that display positive values (e.g. N21).

3.9.3.4 Influence of backbone dynamics

In this section, the possibility that temperature coefficients are related to backbone dynamics is considered. It was hypothesised that amides forming intramolecular

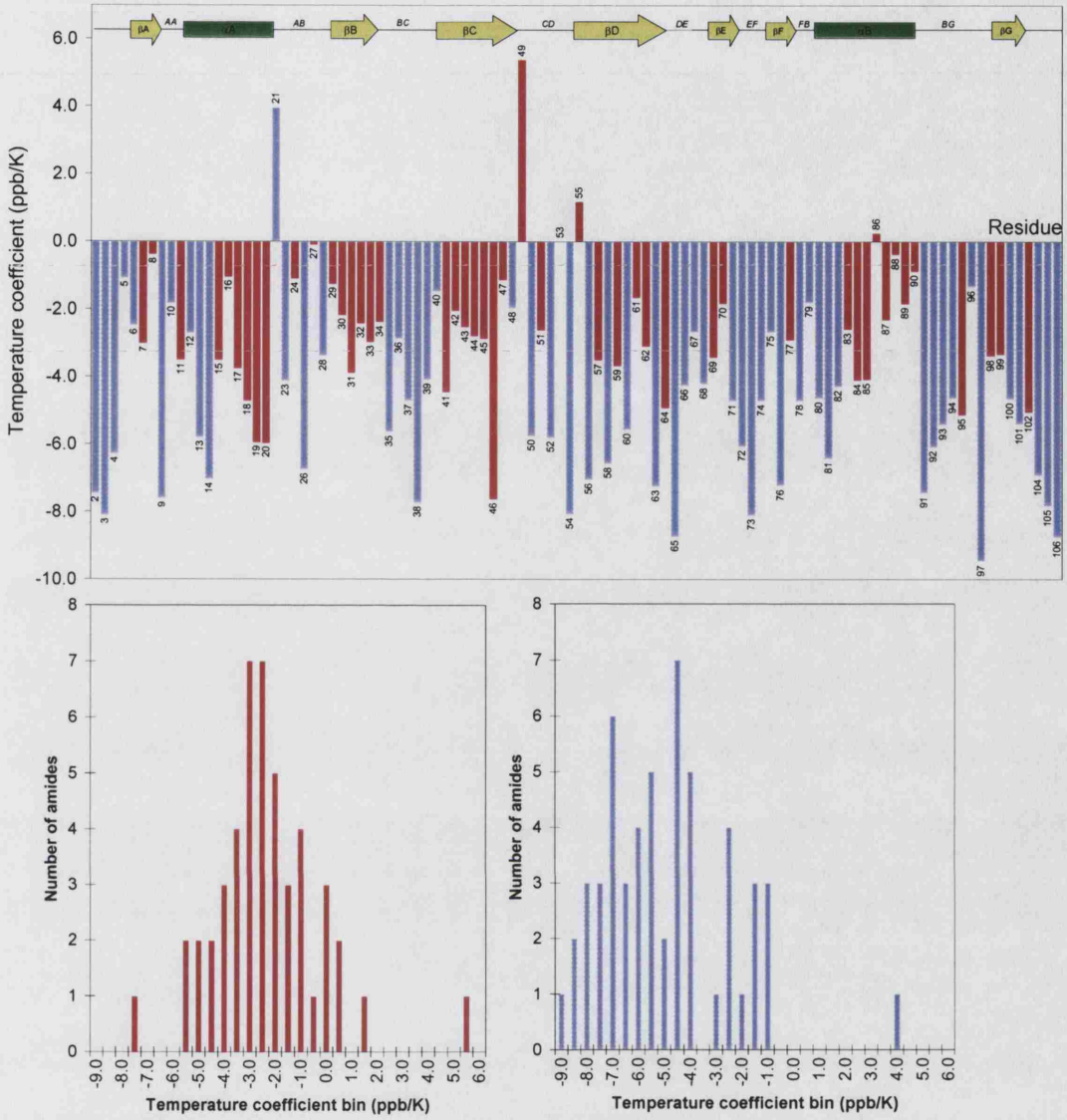


Figure 3.29 Annotated amide proton temperature coefficients. *Top* | Bars coloured according to the presence (red) or absence (blue) of intramolecular hydrogen bonds. Secondary structural elements are also shown as yellow arrows (β-strands) and green rectangles (α-helices). *Bottom* | Histograms showing the distribution of temperature coefficient data according to the presence (*left*) or absence (*right*) of intramolecular hydrogen bonding (same colours as *Top*).

hydrogen bonds will show smaller amplitudes of motion and more positive temperature coefficients than amides exposed to bulk solvent. The backbone dynamics of *v*-Src SH2 are described in Chapter 4. An order parameter S^2 that represents the amplitude of motion of each amide bond vector is calculated through analysis of NMR relaxation data. S^2 values are restricted to within the range zero (complete freedom) to one (rigidity).

The relationship between S^2 and temperature coefficients is shown in Figure 3.30.

		Amide–ring distance	
	Full data	<4 Å	>4 Å
Mean	-4.0	-3.2	-4.2
Mean error	0.27	0.56	0.29
St. Dev.	2.7	2.6	2.6
Minimum	-9.5	-6.4	-9.5
Maximum	5.4	5.4	4.0
n	102	22	80

Table 3.15 Statistical analysis of temperature coefficient dataset according to proximity to aromatic rings.

Regression analysis indicates a very poor correlation between the variables, as indicated by the low R^2 value. An independent statistical analysis of the S^2 data indicated that the average amplitude of motion was not significantly lower for amides forming intramolecular hydrogen bonds (0.88 ± 0.06) than for solvent-exposed amides (0.83 ± 0.09). This leads to the surprising conclusion that hydrogen bonding has little effect on the motional freedom of a given amide or, at the very least, that other contributions (e.g. steric clashes and bond angles) are much more dominant. This explains why no correlation was observed with the temperature dependence data for v-Src SH2.

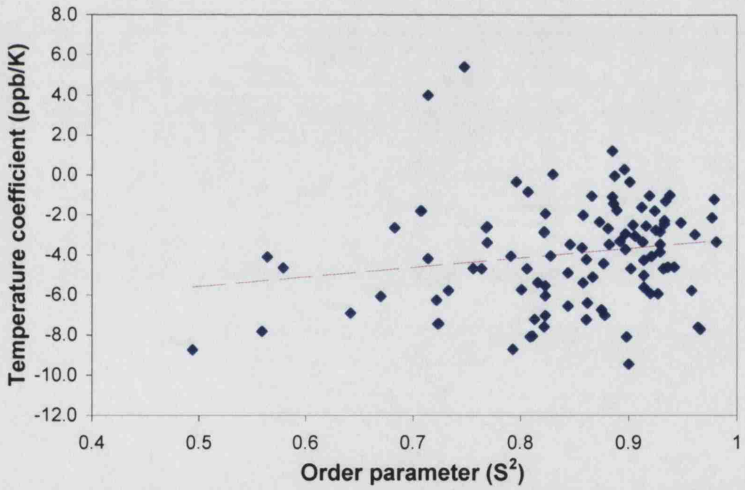


Figure 3.30 Scatterplot showing the relationship between backbone amide dynamics (S^2) and temperature coefficients. Regression analysis within Microsoft Excel generated a best fit line with $R^2 = 0.03$.

3.9.4 Conclusion

v-Src SH2 appears to behave like other small globular proteins and domains in response to temperature. However, the distribution of coefficient data for internal and external hydrogen bonding classes overlapped more than usual, which adversely affected the prediction of hydrogen bonds. The average amino acid chain length of

polypeptides within the published data set was 63, and three proteins consisted of fewer than 30 residues[267]. As *v*-Src SH2 contains 106 residues, it might be expected that different behaviour would be observed.

3.10 Summary of chapter

Our ongoing solution state biophysical studies of the *v*-Src SH2 domain required detailed understanding of the atomic structure in equivalent conditions. This has been achieved through application of standard NMR techniques and the structure determination also provided an excellent opportunity to trial automated NOESY assignment software. The resultant structural ensemble agrees closely with the experimental restraints and is of high quality. As expected, there is strong overall homology between the existing crystal structure and our ensemble of solution structures. Apart from loop regions and termini, the most significant difference between the two structures is at the N-terminal part of the βC strand. It is highly probable that this is caused by the binding of a phosphate ion to the crystal structure. It appears that our ensemble represents the genuine apo state, and is therefore the first of its kind within the Src family of domains, and only the second within the wider super-family of SH2 domains.

Chapter 4

SH2 domain interactions

As discussed in Section 1.3, the SH2 domain provides an excellent test system for the fragment approach to drug design. This chapter describes the effect of ligand binding on *v*-Src SH2 through analysis of three NMR observables – chemical shift, relaxation rate, and signal intensity. These are used to obtain a variety of detailed information regarding the binding event, including changes in electromagnetic environments, backbone dynamics, protonation states, and solvent-accessibility. Chemical shift analysis also provides an indication of ligand affinity. Thus NMR has been exploited in a number of distinct ways to demonstrate its usefulness in the analysis of binding of different SH2 domain ligands as well as enhance our understanding of the target binding site. Information gleaned from these studies may be used to complement or explain thermodynamic binding data as well as guide future ligand design.

4.1 Chemical shift perturbation studies

Interaction between a protein and a ligand results in a specific change in chemical shifts for both molecules. This information may be used as a qualitative indicator of binding or it may be interpreted quantitatively to yield affinity and structural data. The binding interface is located by observing the spatial distribution of affected nuclei. However, an interaction may also lead to changes in structure at distal sites (allosteric mechanism), which will also induce chemical shift changes and thus could be mistaken for a secondary binding site. These secondary effects are interesting in themselves as they allow one to observe the global effect of binding on protein structure.

The aim of this investigation was to explore three different SH2-ligand interactions by monitoring the perturbation of backbone amide chemical shifts from the apo state. The ligands selected were phosphate, phosphotyrosine, and phosphopeptide (PQpYEEIPI), which differ in size, affinity, and the extent to which they interact

with the domain surface. The $^{1}H^{15}N$-HSQC experiment was used to observe NH resonances within the *v*-Src SH2 domain and ligand was titrated until protein sites became fully saturated.

The response of amide resonances to ligand binding depends on the nature of the dynamic binding equilibrium and the properties of the nuclei involved. The interaction between two molecules can be characterised by a chemical exchange rate, R_{ex}, which is related to the on and off rates, k_{on} and k_{off}, respectively. Exchange of a nucleus between two hypothetical electromagnetic environments A and B will result in perturbation of the line shapes and chemical shifts in the NMR spectrum. The actual effect depends on the magnitude of R_{ex} relative to the absolute Larmor frequency difference ($\Delta\omega = |\omega_A - \omega_B|$) between the two states. When the rate of exchange is slow in comparison to $\Delta\omega$, resonance peaks corresponding to A and B are both visible in the spectrum, but are slightly broadened compared to linewidths associated with A and B in the absence of chemical exchange. At the other end of the scale, when $R_{ex} \gg \Delta\omega$ (fast exchange), the apparent Larmor frequency is the mean value of ω_A and ω_B, weighted according to the relative populations of each state, p_A and p_B:

$$\overline{\omega} = \frac{\omega_A p_A + \omega_B p_B}{p_A + p_B} \tag{4.1}$$

Thus a single, coalesced peak is observed at $\overline{\omega}$. At intermediate exchange rates, the effect on the NMR spectrum is more complicated. Upon increasing R_{ex} from slow to intermediate rates, the individual lines gradually broaden and subsequently merge into a single peak, which becomes progressively narrower as R_{ex} approaches the fast exchange limit.

4.1.1 SH2—phosphate binding

The importance of phosphate in conferring high-affinity binding to SH2 domains has been consistently highlighted in drug design efforts[2]. The binding site for phosphate consists of several hydrogen bond donors that can also interact with other small, anionic ligands such as citrate and malonate[277]. Compounds that fail to make this set of interactions are usually associated with a marked loss of affinity[181]. In order to further understand the interaction between phosphate and the *v*-Src SH2 domain, phosphate ions were titrated into ^{15}N-labelled protein, observing the effect on $^{1}H^{15}N$-HSQC crosspeaks. Details of spectral parameters and processing can be found in Sections 2.3.2 and 2.3.3. A 1 M phosphate stock solution (aqueous) was prepared,

using an appropriate mixture of mono- and di-basic phosphate sources*. Phosphate was then added to a 200 μL ^{15}N-labelled sample of *v*-Src SH2 (0.5 mM) to a final concentration of 16.5 mM. This solution was then added in 10 x 8 μL injections to a 600 μL sample of ^{15}N-labelled *v*-Src SH2 (0.5 mM), recording ^{1}H^{15}N-HSQC spectra after each addition, achieving a final phosphate concentration of 1.9 mM – approximately four-fold excess over SH2 domain. By adding a phosphate/protein mixture to the protein sample, the total protein concentration remained constant throughout the titration.

4.1.1.1 Results of phosphate titration

The ^{1}H^{15}N-HSQC spectra obtained at different phosphate concentrations are shown overlaid in Figure 4.1. It is clear that phosphate is in fast exchange with the SH2 domain and that it causes only a small subset of peaks to undergo changes in chemical shift. This suggests that binding is weak, but specific, with a probable dissociation constant in the low mM range.

In order to assess the relative significance of perturbations for each amide, the difference in chemical shift between the apo and phosphate-bound spectrum was quantified in both the ^{1}H and ^{15}N dimensions (see bar graph in Figure 4.1). Since the chemical shift range of ^{15}N nuclei is considerably larger than that of ^{1}H, summation of raw chemical shift differences between the free and bound states will favour amides that experience changes in ^{15}N frequency. A scaling factor is usually applied to ensure that ^{15}N and ^{1}H data can be combined in a more robust manner. However, it is not immediately apparent what this factor should be as ^{15}N and ^{1}H nuclei show differing susceptibilities towards chemical shift perturbation induced by changes in their local electromagnetic environment.

We have scaled ^{15}N data here by using population averaging of over 10^5 chemical shift values as published in the BioMagResBank (BMRB) database†. The standard deviation of ^{15}N and ^{1}H frequencies for backbone amides gives an overall impression of the biologically-relevant range experienced, averaging out the effects of sequence and structure. The ratio of standard deviations for ^{15}N and ^{1}H shifts was found to be approximately 6.4, and this value was used to scale down ^{15}N chemical shift changes within this project. Others have selected integer values between 5 and 10 or used the formula $\Delta\delta_{total} = \sqrt{[\Delta\delta_{^1H} + (\Delta\delta_{^{15}N}/25)]/2}$ [39, 278, 279].

The spatial pattern of phosphate-induced shift changes was revealed by colouring the solution structure of *v*-Src SH2 according to the magnitude of the combined shift

*In order to generate a phosphate solution at pH 6.0, a mixture of 87.7 % $NaHPO_4$ and 12.3 % NaH_2PO_4 by mass was used.

†http://www.bmrb.wisc.edu/

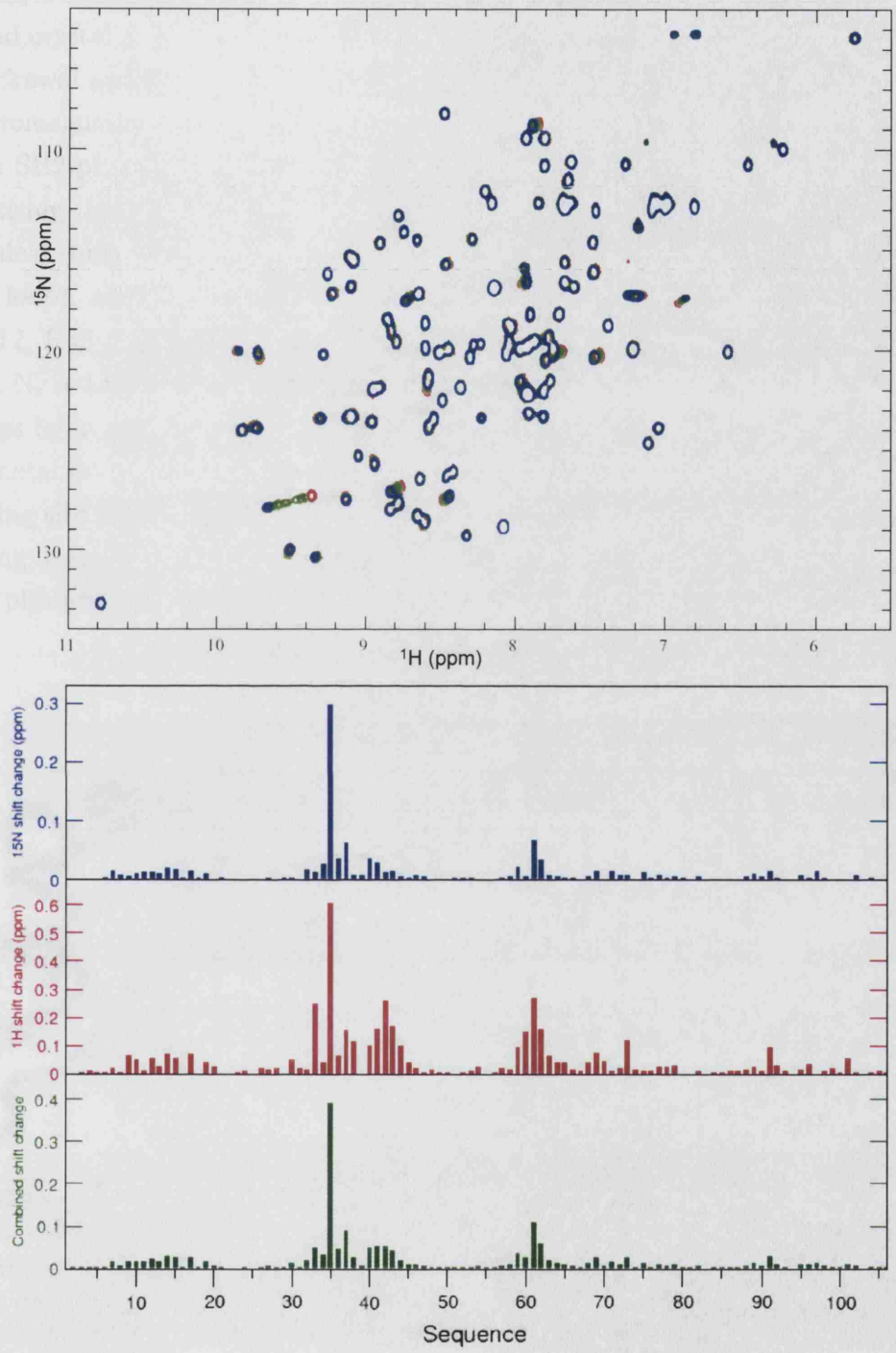


Figure 4.1 Results of phosphate titration. *Top* | Overlaid $^{1}H^{15}N$-HSQC spectra corresponding to apo (red contours), intermediate phosphate concentrations (green), and final concentration (blue). *Bottom* | Chemical shift changes are presented as raw ^{15}N (blue) and ^{1}H (red) data, as well as in a combined form (green).

differences (Figure 4.2). By far the largest perturbation is observed for the amide of E35, which acts as a hydrogen bond donor to a phosphate oxygen in the peptide-bound crystal structure. The strong electronegativity of this oxygen causes electron withdrawal away from the amide proton, rendering it more susceptible to the local electromagnetic field (*deshielding*) and causing an increase in chemical shift. All other SH2-phosphate interactions involve polypeptide side chains, and thus it is not surprising that E35 is the only backbone amide that undergoes a large change in chemical shift. Relatively small perturbations are observed for residues 8-17, 32-42 (BC loop), and 59-62, which reflect interaction of phosphate with the side chains of R12, R32, S34, T36, and C42. The side chain ϵNH resonance of R32 (H_ϵ = 6.9 ppm, N_ϵ = 117.5) also shows sensitivity to phosphate binding as the proximal amine groups hydrogen bond with phosphate. It is somewhat surprising to observe small but notable chemical shift changes for some amides far away from the phosphate binding site (e.g. F69, S73 & A91). Since these residues do not cluster in space, this finding does not provide evidence for an alternative binding site. Instead, interaction with phosphate appears to induce a slight restructuring of the domain at these sites.

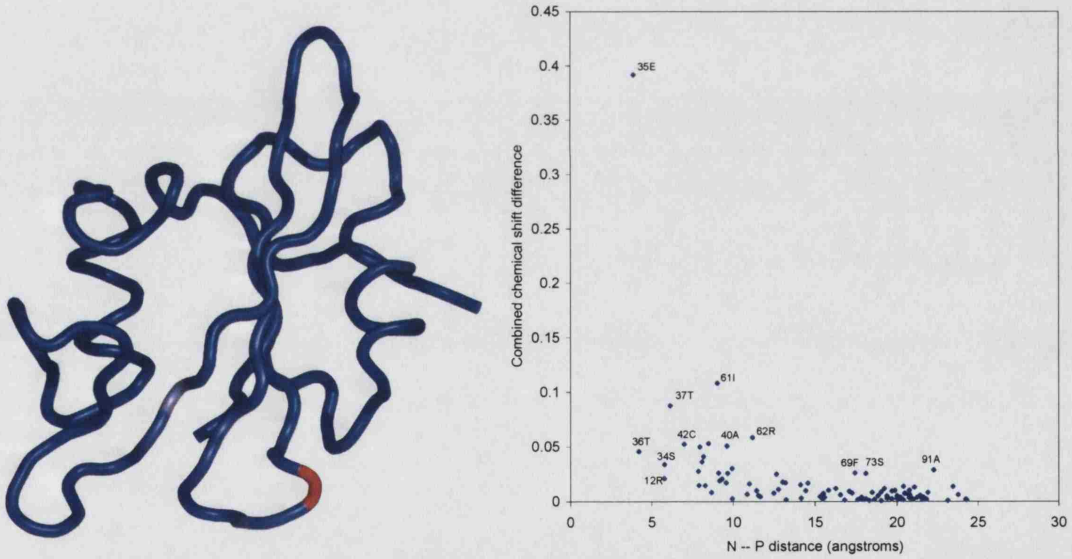


Figure 4.2 Spatial analysis of phosphate-induced chemical shift changes. *Left* | The *v*-Src SH2 solution structure is coloured according to the per-residue combined chemical shift change displayed on binding phosphate. The blue/white/red gradient indicates regions undergoing negligible/small/significant shift changes. *Right* | Scatterplot describing the relationship between ^{15}N–^{31}P distance and chemical shift perturbation.

In order to examine the relationship between ligand proximity and amide chemical shift changes in a more rigorous manner, the distance between each amide nitrogen and the ^{31}P of phosphate was calculated and plotted against the combined shift change (Figure 4.2). This analysis is based on the peptide-bound crystal structure

(1SPS); it is assumed that the phosphate ion occupies a similar position to that observed in 1SPS. The scatter plot shows that amides located more than ~12 Å from the phosphorus tend to show negligible shift changes whilst amides within this cut-off show a diverse range of values. This suggests that amide–ligand distance is only a weak determinant of chemical shift, and thus it would be difficult to convert chemical shift perturbations into distances. For example, the amides of E35 and T36 are both approximately 4 Å from the phosphorus, yet the combined shift of the former is almost 9 times that of the latter because it forms a hydrogen bond with the ligand. Although binding of a phosphate ion is reported within a ~12 Å sphere, this should not be considered as a general 'rule-of-thumb' since every protein and ligand will behave in a unique manner.

In conclusion, the SH2–phosphate interaction is weak, but surprisingly specific given its small size, low complexity, and potential for hydrogen bonding. The single binding site can be readily detected using the $^{1}H^{15}N$-HSQC titration approach and the existence of a hydrogen bond with E35 is strongly supported by the deshielding of this nucleus manifested in the spectrum.

4.1.2 SH2—phosphotyrosine binding

The titration procedure used to examine phosphate binding was repeated using free phosphotyrosine (pY) amino acid. Addition of this ligand led to many changes in amide resonance frequencies and intensities, which were characteristic of intermediate exchange rates (Figure 4.3). This suggests a K_d within the μM $\rightarrow$ mM range, which is consistent with a previous measurement from ITC (K_d = 0.33 mM)[106]. Line-broadening effects caused approximately 15 % of amide resonances to disappear into the background noise during the course of the titration, and reappear elsewhere in the spectrum towards the end of the experiment. Thus it was not possible to track the migration of every amide resonance in the HSQC at the spectral resolution employed. To obtain unambiguous assignments for the pY-bound state, ^{15}N-edited 3D NOESY-HSQC and TOCSY-HSQC spectra were recorded for the complex, allowing side chain 'spin-systems'[‡] and inter-proton NOEs to be compared to that of the apo domain.

Chemical shift differences between the free and pY–bound states were obtained as before and plotted against sequence in their raw and combined forms (Figure 4.3). As with phosphate binding, specific residues, particularly in the range 32-37 and 58-

[‡]'Spin system' refers to the through-bond connectivities observed within certain NMR spectra. For example, the ^{15}N TOCSY-HSQC correlates amide protons to other protons within the same residue, allowing one to distinguish between intra-residue and inter-residue resonances. As each side chain consists of an characteristic set of protons, in ideal circumstances one can confirm the identity of an amide by this method.

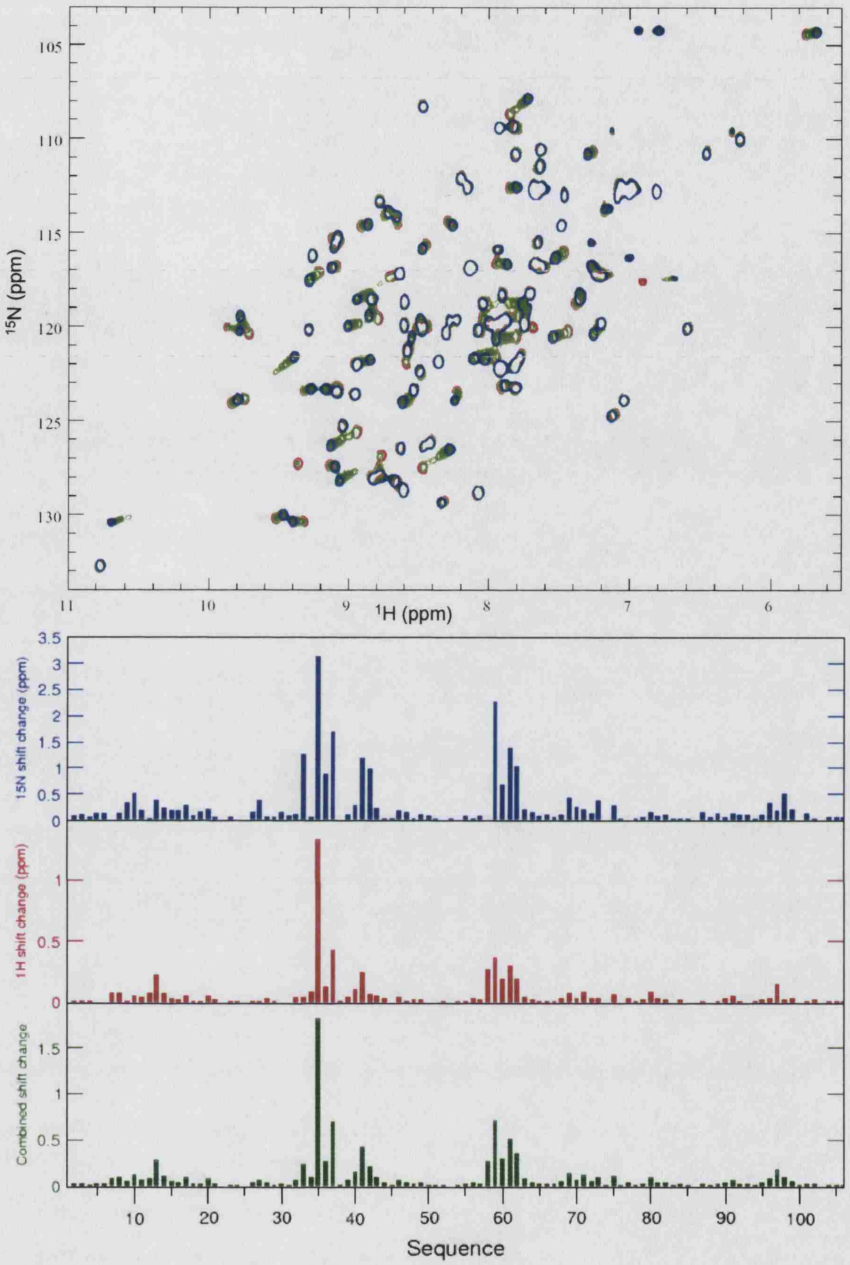


Figure 4.3 Results of phosphotyrosine titration. *Top* | Overlaid $^{1}H^{15}N$-HSQC spectra corresponding to the titration of pY into apo *v*-Src SH2 domain; colours as in Figure 4.1. *Bottom* | Chemical shift differences between free and bound state are depicted as raw ^{15}N (blue) and ^{1}H (red) data, as well as in the combined format (green).

63, display significant perturbations. However, pY appears to exert a more powerful influence on the overall magnitude and extent of shift changes compared to phosphate. This can be explained by two contributing factors: Firstly, pY is known to form more interactions with the domain surface than phosphate in the peptide-bound crystal structure. In particular, the side chains of R12 and K60 form π-interactions

with the aromatic ring. Secondly, due to ring-current effects[§], the aromatic moiety of pY imposes a relatively large electromagnetic field change on many surrounding nuclei. In the presence of phosphate, the amide of E35 was deshielded by 0.3 ppm in the ^{1}H dimension. With the additional aromatic ring of pY, the amide proton is shifted by a further 1.0 ppm. It is possible that the higher affinity of pY involves strengthening of the hydrogen bond with E35 amide, which would also contribute to the deshielding of this nucleus.

The location and magnitude of shift changes induced by pY binding is depicted in Figure 4.4. A high degree of specificity is observed for this ligand; its binding site is readily identified by the clustering of amides undergoing relatively large shift changes. The significant perturbation displayed by H58-R62 within the βD strand indicates that the aromatic group packs against this part of the domain surface, as opposed to against the αA helix. This orientation of pY is also found in the peptide-bound structure of *v*-Src SH2 suggesting that even in its fragment form, the pY ligand adopts a peptide-like binding mode. That a small flexible part of a larger ligand is found to prefer such an orientation indicates that binding is dictated quite strictly by the SH2 domain. Efforts to identify phosphotyrosine mimetics have also resulted in the suggestion that particular binding features and orientations are required for high affinity[181].

The effect of ligand proximity on chemical shift changes was explored again by plotting the distance between each nitrogen and the nearest pY heavy atom against the combined shift change observed for each amide (Figure 4.4). The results show a similar pattern to that observed with phosphate binding, that is, amides located more than 10-12 Å from the nearest ligand heavy atom are usually associated with negligible shift changes. There are exceptions to this, again highlighting F69 as a residue sensitive to binding, yet located over 15 Å from the ligand.

The differential effects of phosphate and pY binding highlight a potential bias in target-based NMR screening towards fragments with inherently large potential to induce chemical shift changes. It is possible that a small, aliphatic, phosphate-based compound devoid of ring-current effects would not induce the same magnitude of shift changes as phosphotyrosine, and therefore this fragment could be passed by irrespective of its affinity.

4.1.3 SH2—PQpYEEIPI binding

Phosphotyrosyl-peptide binding was also investigated using the titration method, although saturation required only a 2-fold excess of peptide over SH2 domain. Figure 4.5 displays the resultant HSQC spectra overlaid. The behaviour of amide

[§]Flow of electrons around a π-orbital system, such as a benzene ring, generates a magnetic field.

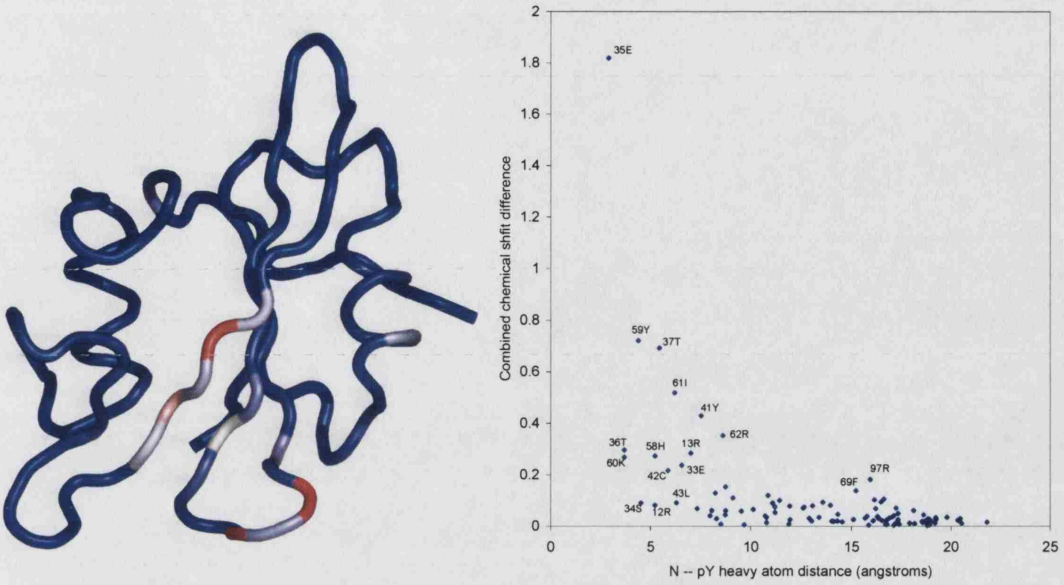


Figure 4.4 Spatial analysis of pY-induced chemical shift changes. *Left* | The apo *v*-Src SH2 structure is coloured according to the magnitude of amide shift changes associated with pY binding; colours as in Figure 4.2. *Right* | Scatterplot depicting the relationship between ^{15}N–pY distance (nearest heavy atom) and the chemical shift perturbation experienced. Distances were calculated on the basis of the peptide-bound *v*-Src SH2 structure.

resonances indicated that the phosphotyrosyl peptide was in slow exchange with the protein. Resonances corresponding to the free and bound states were visible simultaneously during levels of intermediate saturation. Since resonances did not migrate steadily across the spectrum during the titration, bound-state assignments were obtained in the same manner as for the pY-bound complex, using ^{15}N-edited NOESY-HSQC and TOCSY-HSQC spectra.

Chemical shift differences associated with peptide binding were calculated as above and plotted against the domain sequence (Figure 4.5). Interaction with phosphopeptide clearly has a profound effect on the ^{1}H^{15}N-HSQC spectrum. In comparison with phosphate and pY, the peptide ligand engages additional parts of the domain surface, especially amides within the pY+3 pocket (Y70-Q76 and S88-R97). The differences in chemical shift between the apo and peptide-bound states are displayed on the apo structure in Figure 4.6. The largest changes occur, as expected, at the peptide-binding interface. Inspection of the rear surface of the domain indicated an absence of medium to large chemical shift differences. This concurs with previous reports that SH2 domains do not undergo global structural changes on binding peptide[280].

In a similar manner to phosphate and phosphotyrosine ligands, the binding of PQpYEEIPI appears to have a general effect on amide shifts up to 12 Å away (Fig-

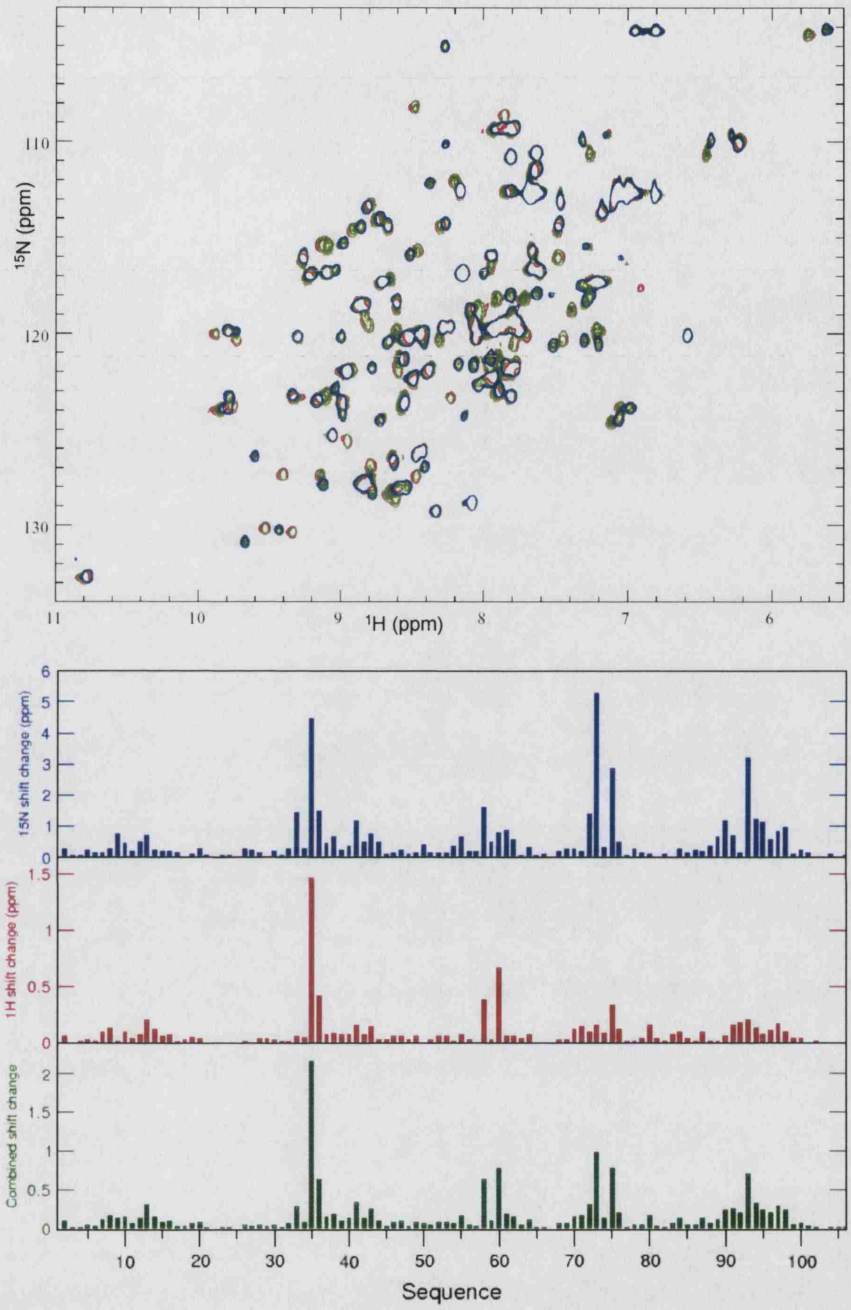


Figure 4.5 Results of PQpYEEIPI titration. *Top* | Overlaid $^{1}H^{15}N$-HSQC spectra corresponding to the titration of PQpYEEIPI into apo *v*-Src SH2 domain; colours as in Figure 4.1. *Bottom* | Chemical shift differences between free and bound state are depicted as raw ^{15}N (blue) and ^{1}H (red) data, as well as in the combined format (green).

ure 4.6). This appears to be equally true for amides in the pY+3 pocket as for those in the pY pocket, suggesting that (in this case) the hydrophobic isoleucine side chain has as much potential to perturb chemical shifts as the aromatic, hydrophilic, and charged phosphotyrosyl moiety. The backbone amide of K60 (Lys βD6) is hydrogen

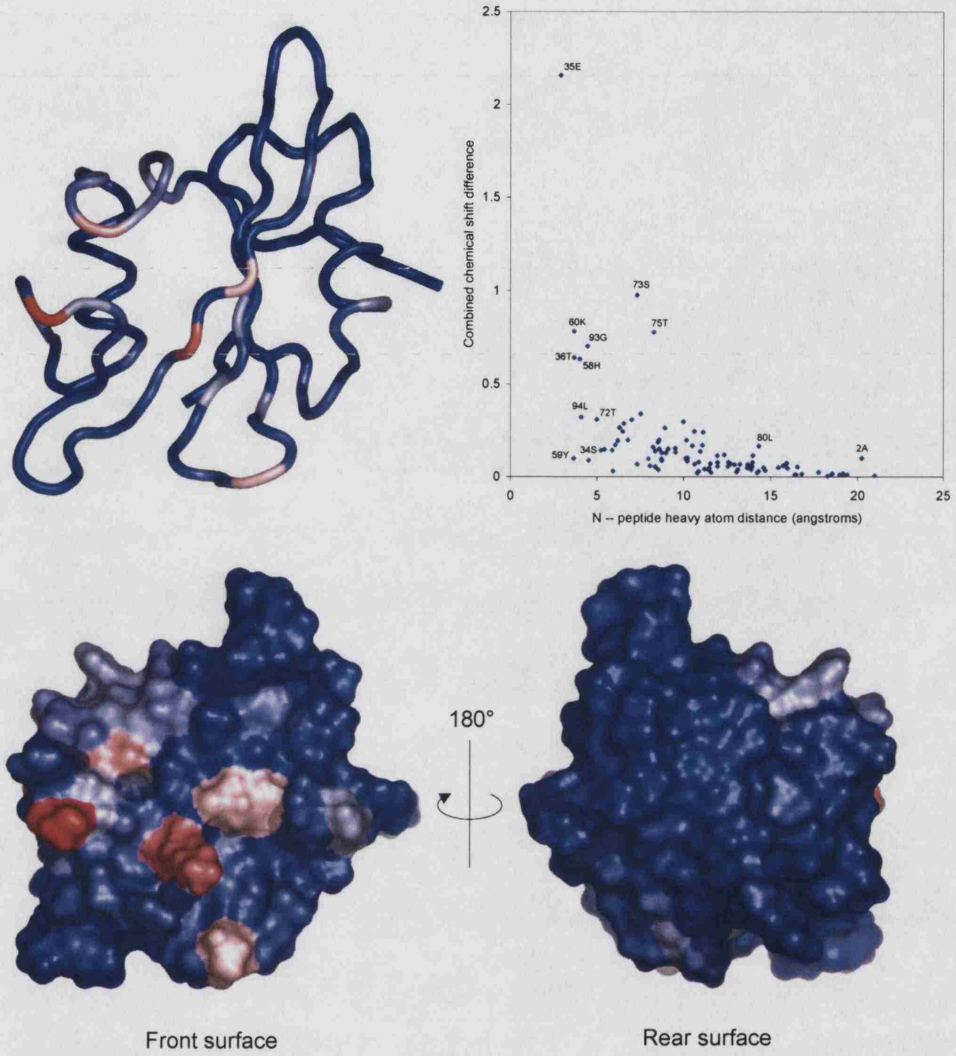


Figure 4.6 Spatial analysis of PQpYEEIPI-induced chemical shift changes. *Top left and bottom* | The apo *v*-Src SH2 structure is shown as a backbone worm and solvent-accessible surface, coloured according to the magnitude of amide shift changes associated with peptide binding; colours as in Figure 4.2. *Top right* | Scatterplot depicting the relationship between the distance separating each amide nitrogen from the ligand (nearest heavy atom) and the chemical shift perturbation it experiences. Distances were calculated on the basis of the peptide-bound *v*-Src SH2 structure.

bonded via water to a backbone carbonyl group of the peptide. This amide reports a combined chemical shift difference between the free and bound states of 0.78. This value is significantly higher than that observed for the pY complex (0.30) in which K60 does not form an important structural hydrogen bond.

4.1.4 Summary of chemical shift perturbation studies

This investigation into the effect of ligand binding on backbone chemical shifts has demonstrated the usefulness of the technique for fragment screening and follow-up analysis. The phosphate, phosphotyrosine, and peptide ligands were found to be in fast, intermediate, and slow exchange regimes with the domain, respectively, and were localised to the site of SH2 ligand interaction. Formation of a hydrogen bond with a backbone amide was usually associated with large chemical shift changes. However, in some cases, interaction of the ligand with a particular side chain (e.g. R32) did not result in a significant shift change for the associated amide of that residue. Thus, shift perturbation data do not indicate the relative importance of individual interactions, they merely report local changes in electromagnetic environment, which may reflect proximity to ligand or structural rearrangement. Overall, the binding of each ligand does not appear to induce structural changes at distal sites, which is entirely consistent with structural data for Src and all other SH2 domains.

The combined chemical shift changes induced by phosphate, pY and PQpYEEIPI binding are compared directly in Figure 4.7. It is clear that phosphate has only a relatively minor effect on chemical shift compared to the larger ligands. The amide of E35 displays the largest perturbation is each case, being deshielded by the direct hydrogen bond and affected by the phosphotyrosyl ring. The additional perturbation observed for E35 in the peptide-bound state may reflect a strengthening of the hydrogen bond compared to the pY-complex. Both pY and peptide induce similar chemical shift changes for the pY pocket and surrounding sites, however it is interesting to note that pY also causes small perturbations for residues in and around the pY+3 pocket (70-100), suggesting that it may modulate the structure of this region. In the context of the cell, pY is never found in this fragment form, rather one must always consider the larger pY-based peptide that interacts with both sites. However, thermodynamic data has highlighted the dominance of pY in terms of binding energy, and it is possible that this part of the peptide binds first, and the C-terminal residues (e.g. EEI in the case of Src) then bind immediately after to strengthen the interaction, although there is no supporting evidence for this. It may be possible that in the context of peptides, phosphotyrosine induces some slight structural or dynamic change in the pY+3 pocket that may have some beneficial effect on the interaction with C-terminal residues. The chemical shift evidence presented here suggest that the electromagnetic environment in this region is slightly altered by pY, however detailed structural and dynamics information is required in order to comment on the validity of this hypothesis.

The distance analyses indicated that ligand binding has a general, decreasing

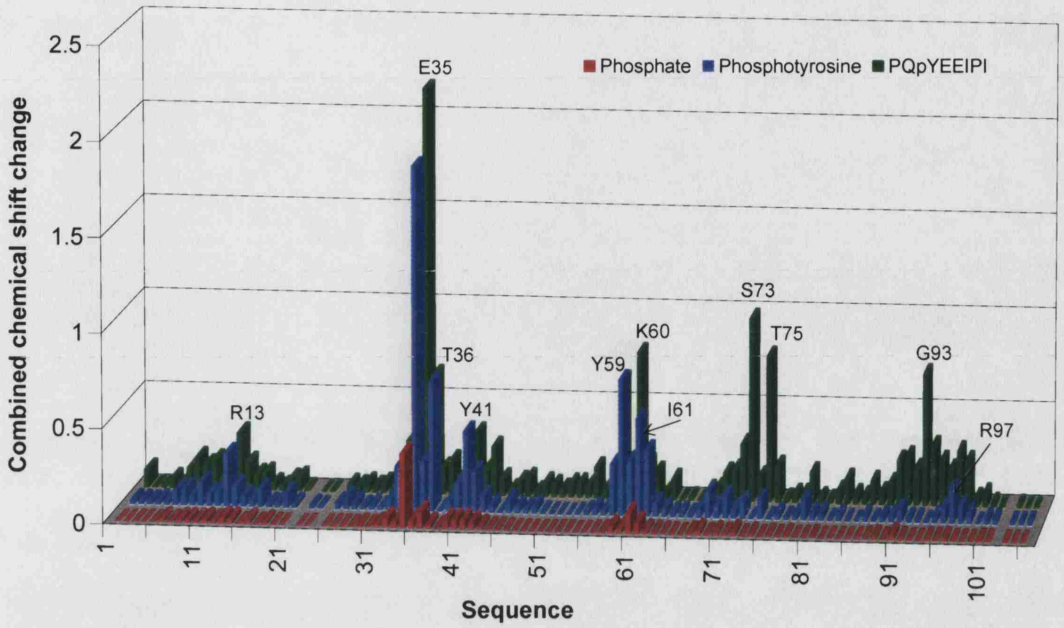


Figure 4.7 Comparison of the combined amide chemical shift changes induced by phosphate, pY, and peptide ligands.

influence on amide chemical shifts up to a threshold of approximately 12 Å. Beyond this distance, shift changes were often negligible, except in a few interesting cases. For amides within the 12 Å sphere, a large variety of shift changes were observed, some of which can be accounted for by invoking hydrogen bonds or ring current effects. This information may be used to locate the binding site, however a more rigorous physical interpretation of amide chemical shift changes would be required to generate of a structural model of the complex.

4.2 Ligand binding and protein backbone dynamics

The chemical shift changes associated with phosphate, pY, and peptide binding indicated parts of the SH2 domain that undergo a change in electromagnetic environment. The effects of interaction with these ligands was further explored by examining the impact of each on the dynamic behaviour of the polypeptide backbone. This information can be obtained by relating NMR signal relaxation rates to a model of the motion of the amide bond vector. Whilst few structural changes are associated with SH2 domain binding interactions, it is possible that the dynamic profile of this domain is affected. Comparison of the dynamic profile of a protein in the absence and presence of a ligand reveals the effect of binding on local and global motion, which may in turn help to understand binding thermodynamics, structural changes, ligand specificity, or biological mechanisms. Hence this approach may provide much needed insight into why one pocket is so specific towards pY and the

other is more tolerant to changes in the ligand.

The dynamic properties of a protein can be related to several different NMR observables, depending on the particular timescales involved. Motions that occur at similar frequencies to that of nuclear spin precession (i.e. MHz–GHz, or ns–ps timescales) are found to largely control relaxation rates. Empirically, therefore, measurement of spin relaxation rates can be used to identify and quantify such motions. There are a number of established procedures for measurement of relaxation rates and interpretation of data. The most common approach involves analysis of ^{15}N-labelled backbone amides.

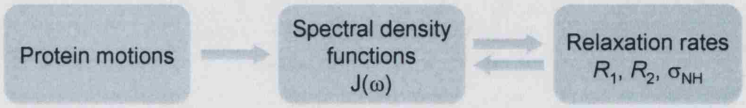


Figure 4.8 Schematic diagram showing the mathematical relationships linking protein motions to observed nuclear relaxation rates *via* the spectral density function.

Molecular motions on the ns–ps timescale and NMR relaxation rates are linked mathematically to a central parameter, termed the *spectral density function*, $J(\omega)$ (Figure 4.8). This function describes the power available at frequency ω to induce fluctuations in the local magnetic field, which ultimately causes relaxation. A ^{15}N nucleus is relaxed by two mechanisms – dipole-dipole (DD) and chemical shift anisotropy (CSA) – which are sensitive to the relative spatial orientation of the attached proton. Longitudinal (R_1), transverse (R_2) and cross-relaxation (σ_{NH}) rates are related to the spectral density function by Equations 4.2-4.4:

$$\begin{aligned} R_1 &= (\mathrm{DD}^2/4)[J(\omega_\mathrm{H}-\omega_\mathrm{N}) + 3J(\omega_\mathrm{N}) \\ &\quad + 6J(\omega_\mathrm{H}+\omega_\mathrm{N})] + \mathrm{CSA}^2 J(\omega_\mathrm{N}) \qquad (4.2) \\ R_2 &= (\mathrm{DD}^2/8)[4J(0) + J(\omega_\mathrm{H}-\omega_\mathrm{N}) + 3J(\omega_\mathrm{N}) \\ &\quad + 6J(\omega_I) + 6J(\omega_\mathrm{H}+\omega_\mathrm{N})] \\ &\quad + (\mathrm{CSA}^2/6)[4J(0) + 3J(\omega_\mathrm{N})] + R_{ex} \qquad (4.3) \\ \sigma_{NH} &= (\mathrm{DD}^2/4)[6J(\omega_\mathrm{H}+\omega_N) - J(\omega_\mathrm{H}-\omega_\mathrm{N})] \qquad (4.4) \\ &\text{where } \mathrm{DD} = (\mu_0 h \gamma_\mathrm{H}\gamma_\mathrm{N}/8\pi^2) r_\mathrm{NH}^{-3} \\ &\text{and } \mathrm{CSA} = \Delta\sigma_\mathrm{N}/\sqrt{3} \end{aligned}$$

$\omega_{\mathrm{H(N)}}$ and $\gamma_{\mathrm{H(N)}}$ are the resonance (Larmor) frequencies and gyromagnetic ratios of the ^{1}H (^{15}N) spins, respectively, μ_0 is the permeability of free space, h is Planck's constant, r_NH is the N–H bond length, and $\Delta\sigma_\mathrm{N}$ is the CSA of the ^{15}N spin. Slower amide vector motion (μs – ms) between distinct environments – i.e. chemical or

conformational exchange – affects R_2 according to the exchange rate R_{ex}.

There are two commonly-used mathematical treatments that relate observed R_1, R_2 and σ_{NH} rates to ^{15}N-^{1}H bond vector motion. By rearranging Equations 4.2-4.4 *spectral density mapping* calculates $J(\omega)$ at five different frequencies (0, ω_N, ω_H, $[\omega_H \pm \omega_N]$), which can be interpreted physically by inversion of each frequency to yield timescales of molecular motion.

In alternative approach, termed *model-free*[281], the stochastic motion of NH bond vectors with time is described by a correlation function, $C_1(t)$:

$$C_1(t) = \left\langle P_2\Big(\boldsymbol{u}(0) \cdot \boldsymbol{u}(t)\Big)\right\rangle \tag{4.5}$$

where P_2 is the second Legendre polynomial and $\boldsymbol{u}(t)$ is the time-dependent NH bond vector. For completely random motion of $\boldsymbol{u}(t)$ and infinitely long t, $C_1(t)$ is zero as there is no correlation between the vector at time 0 and time t. Where bond vector motion is restricted by some energetic barrier*, there will be an asymmetric distribution in vector population. Fourier transformation of $C_1(t)$ produces $J(\omega)$, which defines the Lorentzian shape of this distribution:

$$J(\omega) = \int_{-\infty}^{+\infty} C_I(t)\mathrm{e}^{i\omega t}\mathrm{d}t = 2\int_{0}^{+\infty} C_I(t)\mathrm{e}^{i\omega t}\mathrm{d}t \tag{4.6}$$

For isotropic rotational diffusion Equation 4.6 simplifies to

$$J(\omega) = (2/5)\tau_e/(1+\omega^2\tau_e^2) \tag{4.7}$$

where τ_e is the correlation time of motion, and the time constant for a smooth decay of $C_1(t)$. A central assumption in model-free analysis is the existence of independent, exponentially-decaying correlation functions that describe global molecular tumbling and internal bond vector dynamics. In other words, rotation of the protein and motion of individual bond vectors have no influence over each other. Thus the total correlation function $C(t)$ is the product of C_G, the correlation function for global rotation, and C_I, that for internal vector motion. By separating such motions, with their respective timescales, τ_m for global and τ_e for internal motions, an order parameter (S^2) describing the amplitude of vector motion can be calculated:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right] \tag{4.8}$$

$$\text{where } \tau^{-1} = {\tau_e}^{-1} + {\tau_m}^{-1} \tag{4.9}$$

*For exchange between two states, the motional timescale is determined by the energy barrier (ΔE_a) separating the states, which is defined by the Arrhenius relationship: $\tau^{-1} = Ae^{-\Delta E_a/RT}$

Parameters in Equation 4.8 are fitted to R_1, R_2, and σ_{NH} relaxation rate data using Equations 4.2-4.4, yielding S^2 values for individual bond vectors. The timescale for vector motion τ_e is also obtained, but to a lesser degree of accuracy. Lipari & Szabo's model-free analysis has been supplemented[282] by adjusting Equation 4.8 such that internal motions on two timescales are estimated, with S_f^2 and S_s^2 describing order parameters for fast (τ_f < 10 ps) and slow (τ_s) motions, respectively, where $\tau_m < \tau_s < \tau_f$. Our current implementation of model-free analysis attempts to fit relaxation data to five alternate parameter sets (Table 4.2).

Parameter set	Fitted parameters
1	S^2
2	S^2, τ_m
3	S^2, R_{ex}
4	S^2, τ_m, R_{ex}
5	S_s^2, τ_m, S_f^2

Table 4.1 Lipari-Szabo model free analysis includes five alternative fitting routines

In the first and third routines, it is assumed that $\tau_e \ll \tau_m$. The contribution of R_{ex} to R_2 is ignored for the first two parameter sets. The third and fourth parameter sets include the full Lipari-Szabo analysis whilst the fifth attempts to include the contributions to motion from two different timescales. By definition, S^2 is restricted to the range 0 to 1 where low values indicate unrestricted motional freedom whilst high values are associated with small bond vector motional amplitudes. Thus for regions of secondary structure within proteins, N–H bond vectors often display high (> 0.8) S^2 values. Conversely, loop and terminal regions, tend to produce lower ($0.3 < S^2 < 0.8$) values[283].

4.2.1 Measurement of R_1, R_2, and σ_{NH}

Execution of a ^{15}N relaxation experiment involves creation of non-equilibrium spin states and observation of equilibrium recovery. R_1 can be measured by inverting the equilibrium alignment of spins in the magnetic field, which relax back to the initial distribution at the longitudinal relaxation rate. Net magnetisation can also be rotated perpendicular to the external field, which relaxes back to equilibrium at the transverse rate constant R_2. Heteronuclear NOEs are measured by saturating the ^{1}H nucleus and observing changes in the ^{15}N nucleus, which occurs at the heteronuclear cross-correlation relaxation rate σ_{NH}. Thus, determination of relaxation parameters for ^{1}H-^{15}N spin systems utilises ^{1}H-detected heteronuclear correlation inversion-recovery (R_1), Carr-Purcell-Meiboom-Gill (CMPG, R_2), and steady-state $\{^1$H$\}$-^{15}N NOE experiments[284]. Inversion-recovery and CMPG decay curves are obtained by performing a series of experiments in which the relaxation period T is

varied. In these experiments, the signal intensity $I(t)$ observed at time t is:

$$I(t) = I_\infty - [I_\infty - I_0]\mathrm{e}^{-T/R_1} \tag{4.10}$$

where I_0 and I_∞ are the peak intensities for $T = 0$ and $T = \infty$, respectively. For steady-state NOE experiments, pairs of spectra are recorded with and without ^{1}H saturation and the NOE is given by $I_{\mathrm{sat}}/I_{\mathrm{unsat}}$, where I_{sat} and I_{unsat} represent peak intensities in spectra obtained with and without saturation.

4.2.2 Relaxation data analysis

R_1, R_2 and HNOE relaxation data was collected for the apo *v*-Src SH2 domain as well as for complexes with phosphate, phosphotyrosine, and phosphopeptide (for experimental details, see Section 2.3.2.1). The SH2-ligand samples analysed in the chemical shift perturbation studies described above were used in the relaxation experiments for this chapter. Per-residue R_1, R_2, NOE, and S^2 values were calculated in *Mathematica*† using scripts developed by Drs M. Williams and M. Pfuhl. R_1 and R_2 relaxation rates were obtained by non-linear least squares fitting of the change in peak intensity with relaxation delay to a single exponential decay function (see Figure 4.9 for examples). Estimates of the uncertainty in the fitted parameters were made from the standard deviation of 200 Monte-Carlo fitting attempts. NOE values represent the ratio of peak heights in the saturated and non-saturated experiments.

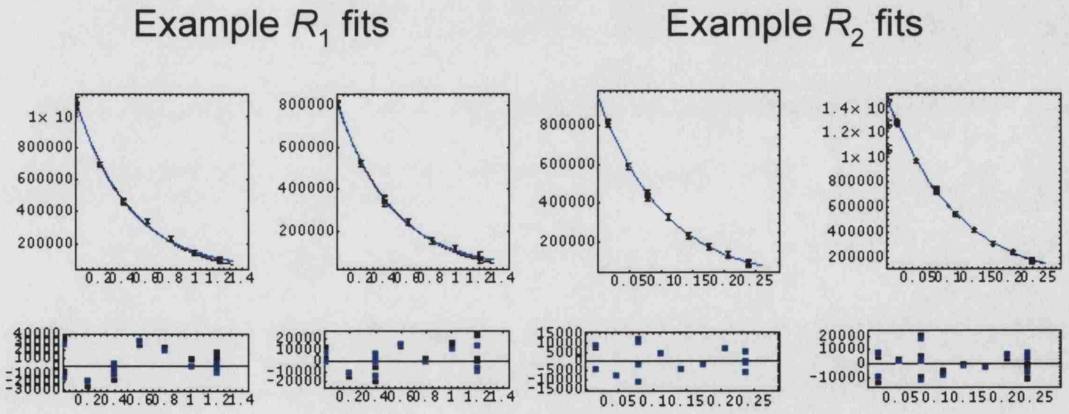


Figure 4.9 Measurement of R_1 and R_2 rates by fitting exponential decay data. The x- and y-axes in the top row of graphs correspond to the experimental relaxation delay and observed signal intensity, respectively. The bottom row of graphs indicates the scatter of data points around the best fit line. Black squares represent empirical data; blue squares indicate back-calculated intensities that would be expected from the fitted rate. In most cases, these match closely the observed data.

The fitted relaxation rates (Table 4.2) were then used to estimate the overall

†*Mathematica 5.0, Wolfram Research Inc.*, Champaign, IL (2003).

tumbling time of the protein (τ_m). Filters[‡] were applied to ensure that residues undergoing large amplitude fast motions or slow chemical exchange and those that represent outliers, were excluded from the prediction. The amide R_2/R_1 ratio provides an indication of local tumbling time, which is compared below to the mean ratio. Since each amide relaxation rate measurement is associated with a variable error (δR_n), the combined error in R_2/R_1 ratio is given by

$$\% \text{ error in } \frac{R_2}{R_1} = 100 \times \sqrt{\frac{\delta {R_2}^2}{{R_2}^2} + \frac{\delta {R_1}^2}{{R_1}^2}}. \tag{4.11}$$

Relaxation data was subsequently fitted to Lipari-Szabo models 1-5, appropriate for the field strength employed (11.74 T) and the calculated value of τ_m. Whilst model-free is the standard analytical approach to interpretation of relaxation data, its usage is controversial[285] since some of the models involve fitting three datasets to three parameters, leading to inaccuracies and imprecision in S^2 values. By including additional relaxation data collected at multiple magnetic field strengths, the ratio of knowns to unknowns can be increased and fitting errors reduced. The Lipari-Szabo method fits amide relaxation data to all five motional models and selects the result that is associated with the smallest statistical error. Occasionally, the 'best' model generates an unrealistic S^2 value (>1), suggesting that model selection is also a point of weakness in this approach. The relaxation data analysed in this study corresponds to a single field strength and was found to produce unrealistic S^2 values for between 10-30 % of amides. The S^2 data presented should therefore be interpreted with caution.

4.2.3 Results for apo v-Src SH2

Raw R_1, R_2, and ^{1}H-^{15}N NOE relaxation data and S^2 estimates for the apo domain are shown in Figure 4.10. R_1 values are relatively stable across the SH2 domain sequence, except for loop regions, which display highly localised increases in relaxation rate. However, the AB loop appears to behave like ordered secondary structure, as evidenced by R_1 values in the region of 2.1 s^{-1}. This loop contains two proline residues that restrict local motion. The distribution of R_2 data also indicates a correlation with secondary structure, however in this case a decreased relaxation rate is observed for loops as well as the N- and C-terminus. The trend is even more apparent with ^{1}H-^{15}N NOE data, which clearly distinguishes between mobile regions, such as the BC, CD and DE loops, and more ordered sequences forming regular secondary structure. The R_2/R_1 ratio plots indicate that α-helices and β-strands are

[‡]Amides displaying NOE ratios below 0.65, or R_2/R_1 ratios in excess of 1.5× standard deviations from the mean ratio were excluded.

	Apo domain	$+PO_4^{2-}$	+pY	+PQpYEEIPI
Raw data				
$\langle R_1 \rangle$ (s^{-1})	2.33 ± 0.43	2.72 ± 0.85	2.61 ± 0.63	2.33 ± 0.47
$\langle R_2 \rangle$ (s^{-1})	8.68 ± 1.44	9.26 ± 1.28	9.56 ± 1.72	9.95 ± 1.61
⟨NOE⟩	0.69 ± 0.16	0.68 ± 0.18	0.69 ± 0.17	0.68 ± 0.19
$\langle R_2/R_1 \rangle$ ratio	3.81	3.60	3.81	4.37
τ_c fitting				
No. amides used	70	67	76	62
$\langle R_1 \rangle$ (s^{-1})	2.25 ± 0.22	2.54 ± 0.39	2.58 ± 0.49	2.25 ± 0.15
$\langle R_2 \rangle$ (s^{-1})	9.25 ± 0.50	9.62 ± 0.76	9.87 ± 0.92	10.4 ± 0.70
⟨NOE⟩	0.76 ± 0.04	0.75 ± 0.05	0.75 ± 0.05	0.77 ± 0.08
$\langle R_2/R_1 \rangle$ ratio	4.15 ± 0.38	3.86 ± 0.53	3.93 ± 0.66	4.64 ± 0.35
τ_c (ns)	6.74 ± 0.18	6.39 ± 0.26	6.44 ± 0.32	7.26 ± 0.14
Model-free analysis				
model 1	45	53	51	60
model 2	15	22	18	3
model 3	2	3	3	2
model 4	17	15	23	20
model 5	23	9	7	17
$\langle S^2 \rangle$	0.86 ± 0.16	0.89 ± 0.26	0.93 ± 0.17	0.92 ± 0.17

Table 4.2 Summary of relaxation analysis data for the apo *v*-Src SH2 domain, and its complexes with phosphate, phosphotyrosine, and PQpYEEIPI. Raw data describes the mean ± S.D. obtained from fitted relaxation rates and NOE ratios. The τ_c fitting section reports the number of amides used during the process along with mean ± S.D. relaxation data for the refined data set. The S^2 section indicates the total number of residues fitted to each motional model as well as the mean ± S.D. in S^2 values.

associated with motional timescales that are slower than the mean whilst loop and terminal regions undergo structural fluctuations at a faster rate. The information on NH bond vector amplitude, described by S^2, concurs with the relaxation data for the apo domain.

The pattern of data describing the backbone dynamics in apo *v*-Src SH2 is broadly explained by known structural information. It was expected that amides in loops and termini would show fluctuations in excess of secondary structural regions and there does not appear to be anything unusual about the backbone dynamics of *v*-Src SH2 in the apo state. The data suggests that the DE loop is slightly more mobile than other loops.

The structure of *v*-Src SH2, described in the previous chapter, was defined with varying degrees of local precision depending on the number of unique restraints and the absence or presence of medium to long range NOEs (Figure 3.16 on page 126). Comparison of the per-residue ensemble rms deviation from the mean backbone structure to R_2/R_1 ratios indicates that local precision is also related to dynamics (Figure 4.11). Regions that are less well-defined also display below-average R_2/R_1 ratios, suggesting that they undergo more extensive structural rearrangements than other regions with low structural variability. This correlation would only be expected

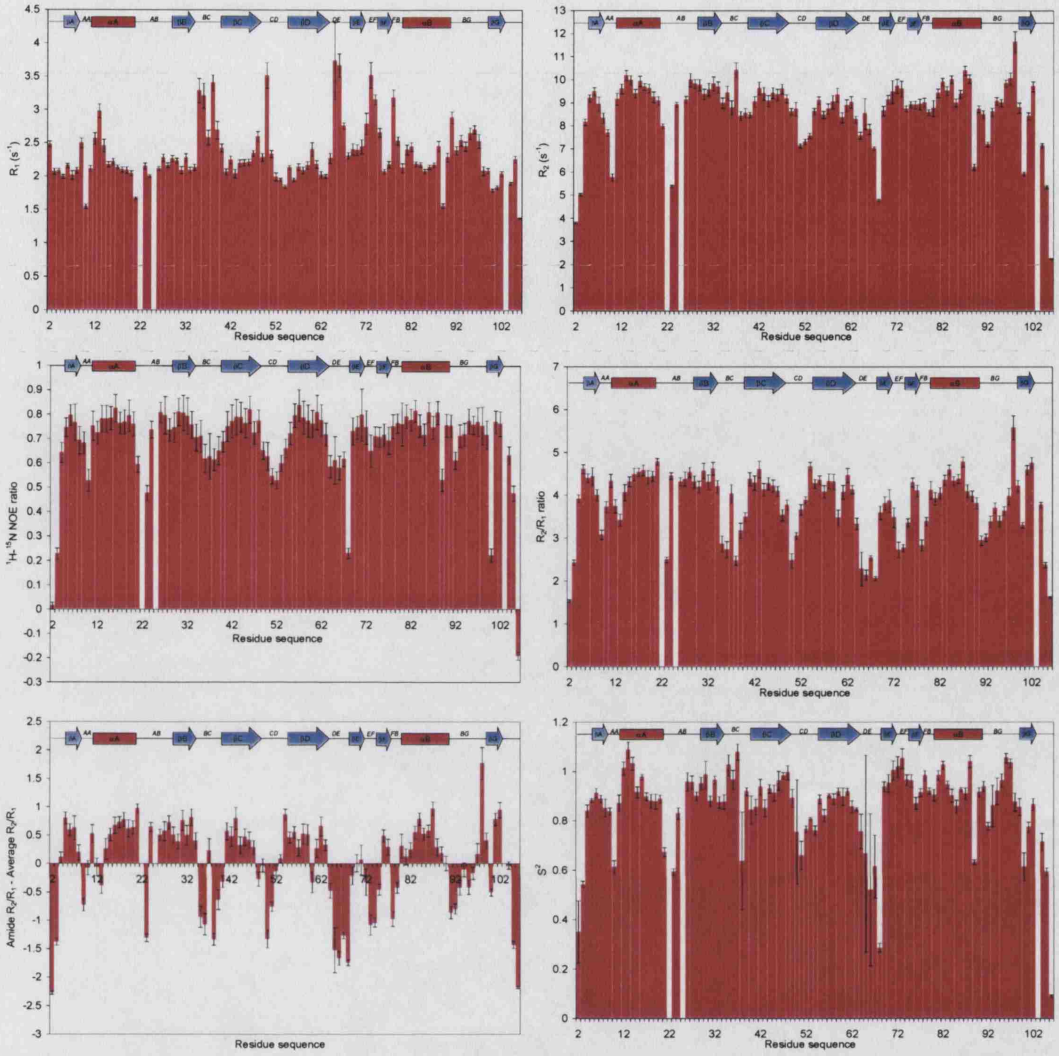


Figure 4.10 Relaxation data for backbone amides within apo *v*-Src SH2. *Top row* | Fitted R_1 (*left*) and R_2 (*right*) values. *Middle row* | ^{1}H-^{15}N NOE ratios (*left*) and R_2/R_1 ratios (*right*). *Bottom row* | Deviation of the amide R_2/R_1 ratio from the average ratio (*left*); S^2 values (*right*). Error bars for raw data graphs represent the errors in fitting. These errors are combined in R_2/R_1 plots in the manner described above (Equation 4.11). Error bars within the S^2 plot represent fitting errors for the selected motional model. The location of α-helices (red boxes) and β-strands (blue arrows) in sequence is indicated at the top of each graph.

if the solution structure is correctly defined.

In comparison to dynamics data reported for the C-terminal SH2 (C-SH2) domain of p85α[286] and for the SAP SH2 domain[287], the relaxation rates of apo *v*-Src SH2 appear to show a stronger dependence on secondary structure. All studies suggest that the BC and DE loops are particularly flexible and p85 C-SH2 displayed significant R_{ex} contributions to R_2 indicating the presence of slow conformational exchange of the BC loop. The τ_c value (6.7 $\pm$ 0.2 ns) obtained prior to S^2 fitting is

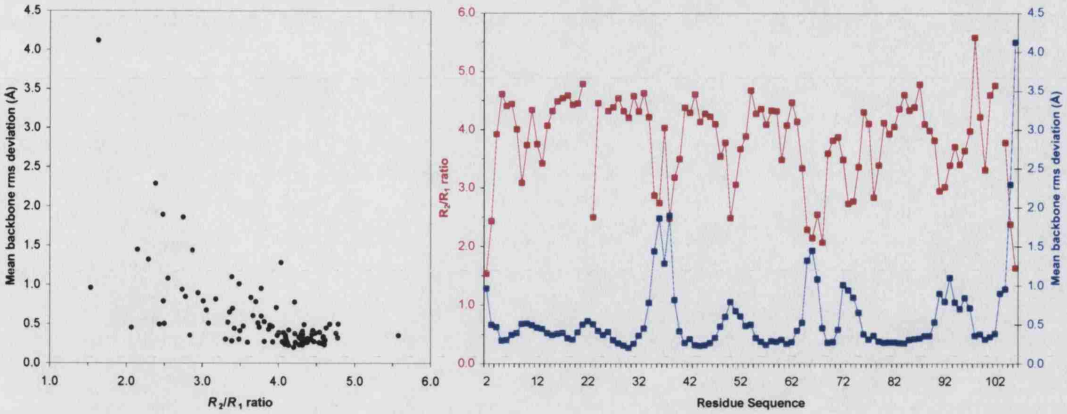


Figure 4.11 Correlation between backbone amide dynamics (R_2/R_1 ratios) and structural precision of apo *v*-Src SH2 backbone atoms. *Left* | Scatterplot comparing R_2/R_1 ratios with mean backbone rms deviations within the structural ensemble. *Right* | Equivalent data displayed for each residue in the polypeptide sequence.

similar to those reported for the SH2 domains of SAP[287] (5.53 ns), Hck[288] (6.8 ns), and PLCγ_1[289] (6.6 ns), as well as other values measured for *v*-Src SH2 using 600 and 800 MHz spectrometers (R. Fawaz, unpublished data).

4.2.4 Results for *v*-Src SH2–phosphate complex

Binding of phosphate to *v*-Src SH2 should have a negligible effect on the overall molecular weight and rotational diffusion coefficient, however it may affect dynamics in and around the binding site. Relaxation data describing this complex yielded a τ_c value of 6.4 ± 0.3 ns, which is within error of that of the apo domain (Table 4.2). The effects of phosphate binding on backbone amide dynamics are shown in Figure 4.12. As described above (Section 4.1.1), phosphate interacts exclusively with the phosphotyrosine binding site. Structural analysis of the peptide-bound state indicates direct hydrogen bonding with the side chains of R12, R32, S34, T36, and C42, and the backbone amide of E35. Residues 37-40 of the BC loop do not make any direct interactions with phosphate or phosphopeptide. NOE data for the phosphate complex suggests a slight decrease in motional freedom for the AA and BC loops compared to the apo domain (Figure 4.12). Other parts of the protein structure do not appear to undergo significant change in NOE behaviour. Addition of phosphate causes significant increases in R_1 for T36 and T37, however differences in R_2 across all amides was slight. S^2 data indicates a general increase in order throughout the domain, including the BC loop, however the large error bars prevent confident detection of changes in dynamics.

Binding of a phosphate ion appears to have a marginal effect on dynamics on the ps–ns timescale that is localised near the binding site. Essentially, the BC loop

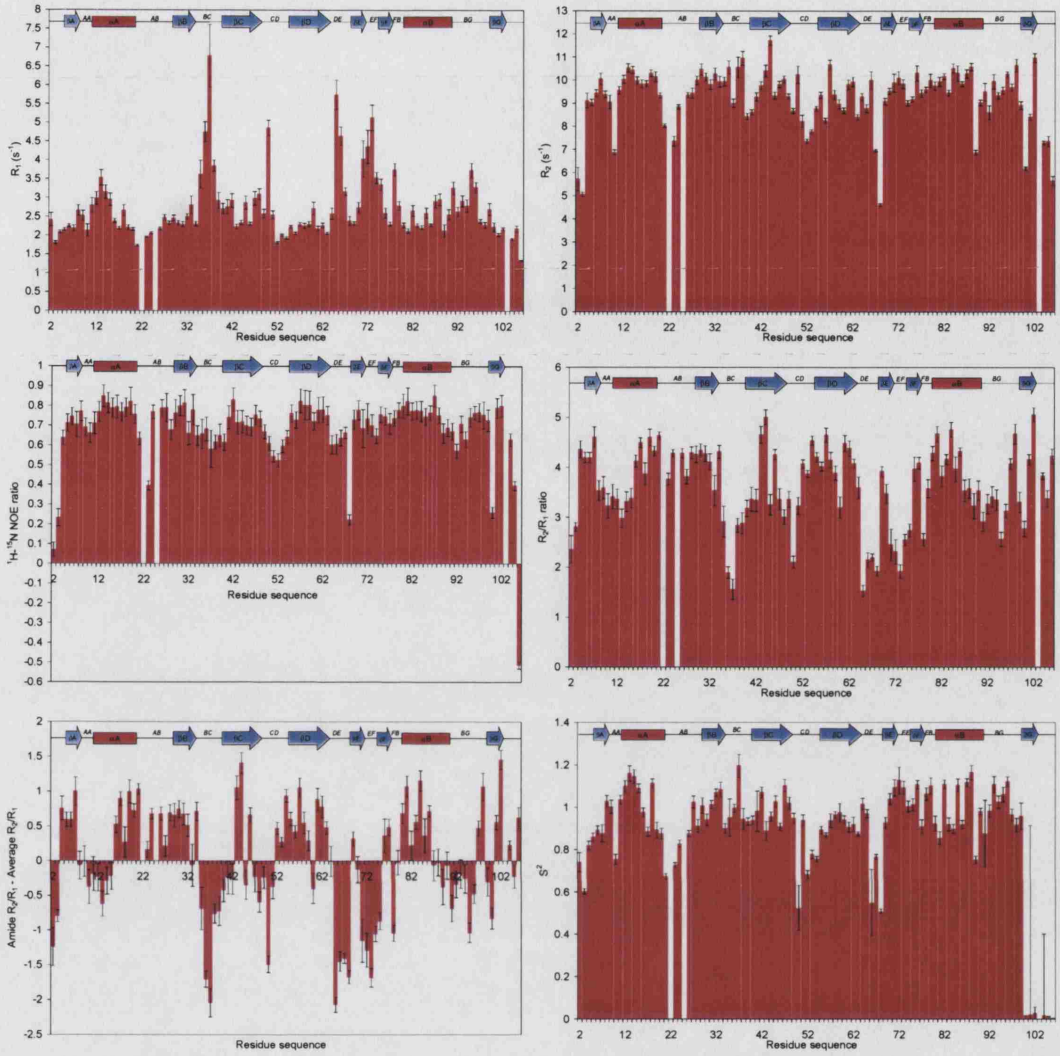


Figure 4.12 Relaxation data for backbone amides within the *v*-Src SH2–phosphate complex. *Top row* | Fitted R_1 (*left*) and R_2 (*right*) values. *Middle row* | ^{1}H-^{15}N NOE ratios (*left*) and R_2/R_1 ratios (*right*). *Bottom row* | Deviation of the amide R_2/R_1 ratio from the average ratio (*left*); S^2 values (*right*). Error bars reflect fitting errors as described in Figure 4.10.

undergoes a slight reduction in mobility. This result may be expected given that the structure of *v*-Src SH2 is well-defined in the apo state and that it does not undergo large conformational changes on binding peptide. Moreover, most of the interactions between phosphate and the domain surface involve side chains rather than backbone amides. In other studies, comparison of protein relaxation and structural data in the absence and presence of metal ions or small ligands, has revealed larger changes in dynamics[290, 291]. Evidence from the analysis of phosphate binding in terms of backbone amide chemical shifts and dynamics confirms that *v*-Src SH2 is capable of recognising this ligand without significant re-ordering of side chains or alteration in

dynamics, both of which contribute towards the relatively large amounts of binding free energy attributed to phosphate within the context of phosphotyrosyl peptides.

4.2.5 Results for *v*-Src SH2–phosphotyrosine complex

Since pY is somewhat larger than phosphate and interacts with a greater proportion of the SH2 domain surface, it might therefore be expected to modulate backbone dynamics to a greater extent. The calculated rotational correlation time for the pY complex is close to that of the isolated domain (see Table 4.2), although the mean errors in R_1 and R_2 fitting were slightly higher. Relaxation data (Figure 4.13) indicates that phosphotyrosine binding also leads to a reduction in mobility for the BC loop backbone amides compared to the apo state; this is particularly evident in the NOE data. The S^2 data also show an increase in order for this region, however the values in excess of 1.0 detract from the validity of this comparison. The similar results obtained the pY and phosphate complexes indicate that in terms of backbone dynamics there is little to distinguish these ligands.

4.2.6 Results for *v*-Src SH2–PQpYEEIPI complex

Of the three SH2 domain ligands, phosphopeptide causes the largest change in the dynamic profile of the apo state. The additional molecular weight of the peptide (~1 kDa) leads to a slight increase in τ_c (7.3 ± 0.2 ns). Residues in the BC loop again display reduced flexibility in the bound state. There does not appear to be much difference in the dynamics of pY pocket backbone amides compared to the analogous complexes with phosphate and pY. Heteronuclear NOE and R_2/R_1 ratio patterns indicate that the CD and DE loops are largely unaffected by interaction with peptide. This is expected from the known crystal structure. Conversely, the short EF and FB loops, and the longer BG loop that form large parts of the pY+3 binding pocket all display a change in relaxation behaviour that is consistent with a decrease in backbone flexibility.

4.2.7 Summary of relaxation results

This investigation into the effect of phosphate, pY (both fragment ligands), and phosphopeptide binding on backbone dynamics has highlighted several important aspects of SH2 domain function. The first conclusion to be made from these data is that in the apo state, the pY pocket is more rigidly defined than the pY+3 pocket. The effects of binding of fragment ligands at the former site are restricted to part of the BC loop. These findings are consistent with the chemical shift perturbation studies which also reported relatively minor effects for sites distal to the pY pocket.

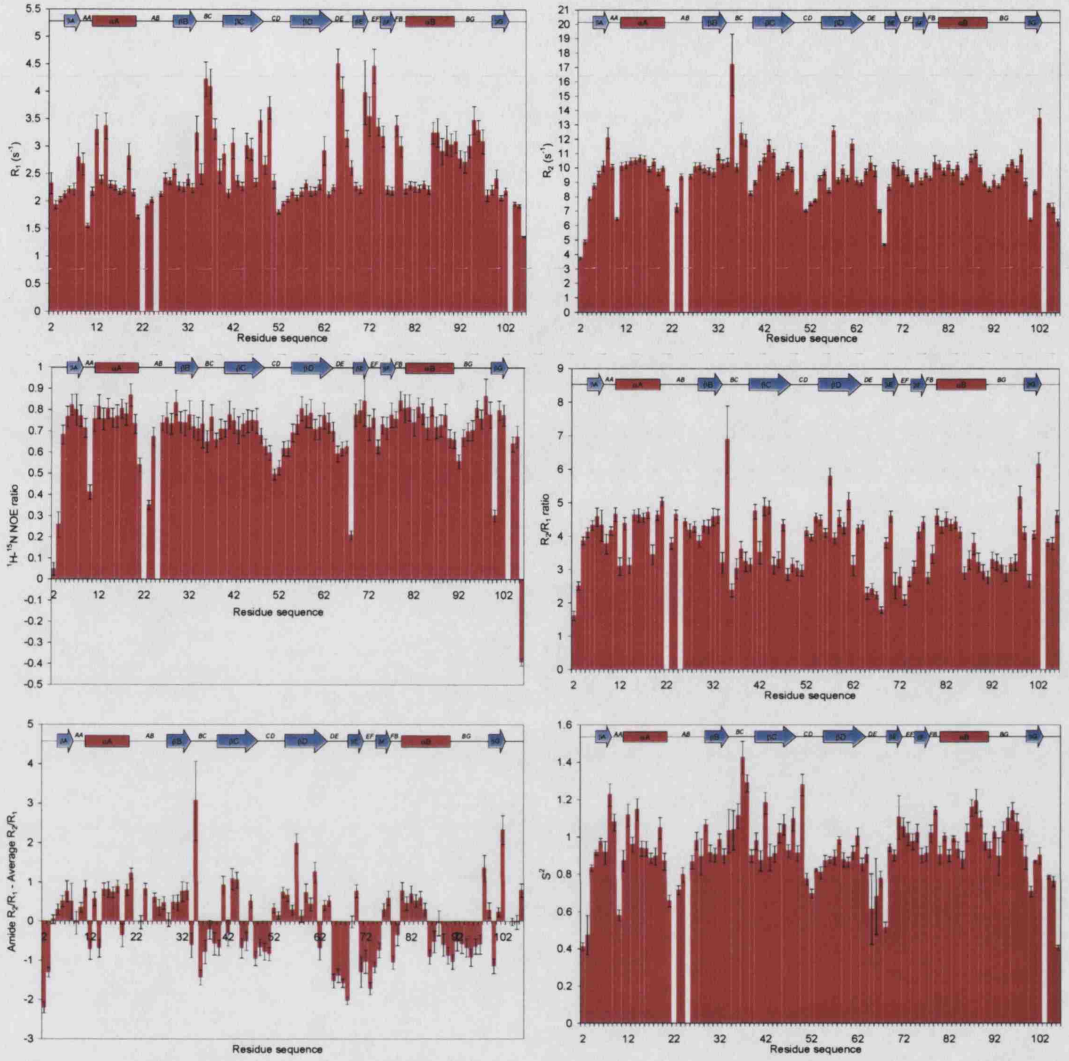


Figure 4.13 Relaxation data for backbone amides within the *v*-Src SH2–pY complex. *Top row* | Fitted R_1 (*left*) and R_2 (*right*) values. *Middle row* | ^{1}H-^{15}N NOE ratios (*left*) and R_2/R_1 ratios (*right*). *Bottom row* | Deviation of the amide R_2/R_1 ratio from the average ratio (*left*); S^2 values (*right*). Error bars reflect fitting errors as described in Figure 4.10.

There is little evidence from these dynamics studies for allosteric changes in SH2 domain structure on binding these fragments.

The difference in behaviour of the pY and pY+3 pockets is readily explained by considering their structure. With the exception of T36, the important side chains that interact with the phosphate and aromatic moieties of pY are located in the αA helix, and βB, βC, and βD strands. This prevalence of regular secondary structure reduces the scope for motion and results in a well-defined binding site. Moreover, it is interesting to note that several of the groups that contact pY are relatively small in size (S34, E35 (amide), T36, & C42) and that the most important contact, the

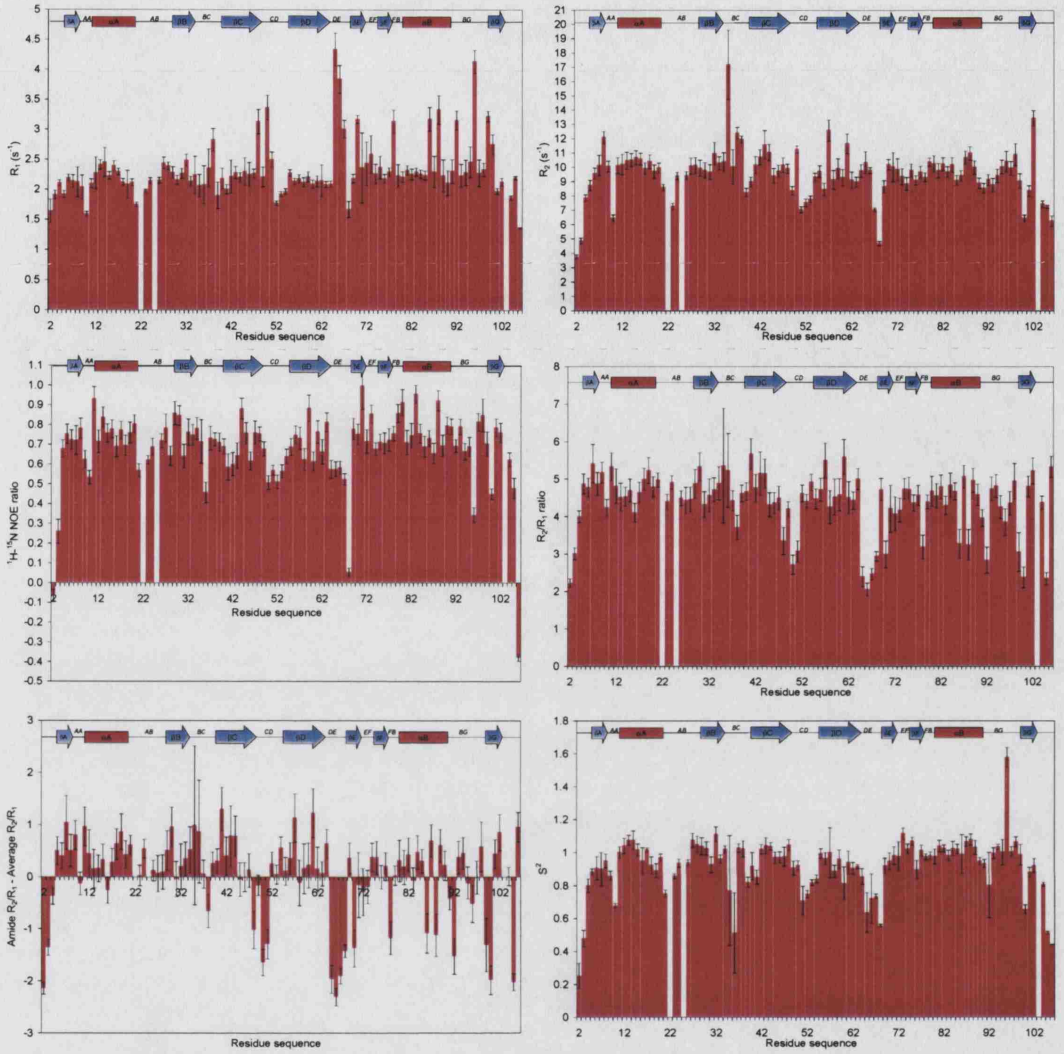


Figure 4.14 Relaxation data for backbone amides within the *v*-Src SH2–PQpYEEIPI complex. *Top row* | Fitted R_1 (*left*) and R_2 (*right*) values. *Middle row* | ^{1}H-^{15}N NOE ratios (*left*) and R_2/R_1 ratios (*right*). *Bottom row* | Deviation of the amide R_2/R_1 ratio from the average ratio (*left*); S^2 values (*right*). Error bars reflect fitting errors as described in Figure 4.10.

guanidinium moiety of R32, is restricted by its high degree of burial and formation of a stable hydrogen bond with H58 imidazole (see below). Thus, apart from R12 and K60 which are relatively long and flexible, there appears to be little adaptive potential in the pY binding site towards diversity in ligand structure.

The structure of the pY+3 pocket is clearly different in that it is formed by the cleft between the EF and BG loops. NMR relaxation data indicates that the flexibility of these regions is above average for the domain in the apo state and is essentially unchanged in the phosphate- and pY-bound complexes. Insertion of the isoleucine of the peptide sequence between these two loops results in a decrease in

flexibility. The high degree of flexibility and hydrophobic nature of the pY+3 pocket is consistent with observations that it will interact with a diverse range of ligands. The adaptability of this site makes it difficult to engineer specificity into a potential drug compound.

Whilst understanding of the SH2 domain backbone dynamics is useful, an equivalent analysis of side chain groups will yield further insights. Unfortunately, the pY pocket of *v*-Src SH2 contains only two methyl groups (T36 & T37), thus the standard ^{13}C-labelled approach may not work in this case. Others have successfully performed relaxation analysis using ^{15}N-labelled arginine side chains in the C-terminal SH2 domain of PLCγ_1[292].

4.3 SH2 side chain ionisation states

This section is focussed upon the role of ionisation events in SH2-ligand interactions. NMR was used to obtain pK_a values for particular titratable groups that are important to binding at the pY pocket. This study was prompted by knowledge that the affinity of *v*-Src SH2 for peptide is dependent on solution pH[193] and because the altered pK_a of phosphotyrosine mimetics, particularly phosphonates, has sometimes been invoked to explain reductions in binding affinity[176]. Knowledge of the protonation states of both protein and ligand can be extremely useful in achieving correct definition of atomic structures for virtual screening or ligand design.

The equilibrium protonation state of an ionisable group is dependent on energetic and kinetic factors such as heteroatom electronegativity (i.e. its 'affinity' for the hydrogen electron), molecular structure, local electrostatic environment, solvent accessibility, and sample conditions (dielectric constant and temperature). Reversible protonation can be described using the generic formula $HA \rightleftharpoons A^- + H^+$, in which A represents an acid. The pH at which equal concentrations of HA and A^- occur is known as the pK_a value, which can be used to calculate the ionisation state of the acid at a given pH*. Empirical values have been determined for amino acid residues exposed to solvent†, however in certain cases where a particular ionisation state is important to the structure or function of a protein, pK_a values deviate from the norm (see Harris & Turner[294] for a recent review).

4.3.1 Literature describing SH2 domain protonation states

Crystalline and solution state structures of several SH2-peptide complexes led Singer & Forman-Kay to investigate the ionisation states of residues implicated in peptide

*This calculation uses the Henderson-Hasselbalch equation: $pH = pK_a - \log_{10} \frac{[HA]}{[A^-][H^+]}$

†For example see Stryer[293]

binding[256]. Histidine pK_a values were obtained for the PLCγ_1 C-terminal SH2 (C-SH2) domain peptide complex by observing changes in $^{15}N_{\delta 1}$ and $^{15}N_{\epsilon 2}$ nuclear resonances with pH. Mutagenesis studies demonstrated that His βD4 plays an important role in SH2 function[295, 296]. His βD4 does not bind phosphotyrosine directly, instead it appears to form a hydrogen bond with the side chain of highly conserved Arg βB5. It was found that resonances for His βD4 were unaffected by pH over the range 5-8, whilst resonances for the other four histidines showed sigmoidal changes indicative of exposure to solvent (pK_a = 6.2 ± 0.8). Moreover the chemical shift value for His βD4 suggested the exclusive presence of the $N_{\epsilon 2}$-protonated tautomer. It was concluded that His βD4 must have a pK_a below 4, which the authors explained by a combination of hydrogen bonding network (Figure 4.15), close proximity to positively charged groups, and solvent inaccessibility.

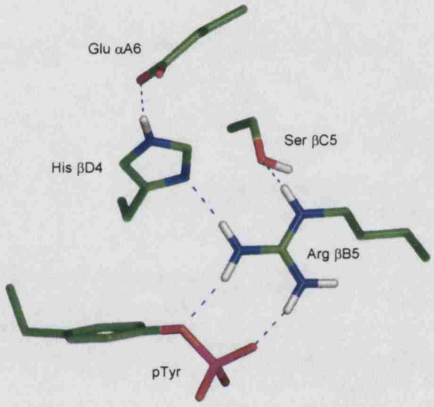


Figure 4.15 A hydrogen bonding network is observed in several atomic structures of SH2 domain complexes. Although Singer & Forman-Kay predict that it exists in PLCγ_1, the published structure (PDB 2PLE) does not include this arrangement of side chains. In fact, the pY binding site in this structure is poorly conserved within the NMR ensemble. Various modes of phosphotyrosine side chain are observed and the crucial Arg βB5 adopts a number of unrealistic orientations. The network is clearer in the crystal structure of Syp N-terminal SH2 domain complex (PDB 1AYA[297]), which is shown in the figure above. Protons were added with PyMOL.

In order to determine whether the pK_a of His βD4 was also below 4.0 in the apo domain (and hence suggest the existence of the hydrogen bonding network) Singer & Forman-Kay compared histidine $^{15}N_{\delta/\epsilon}$ resonances for both the apo and bound states at pH 5.5. The spectra were very similar, indicating that the N_ϵ-protonated tautomer was still preferred, and that the electrostatic environment surrounding each histidine was essentially the same in the free and bound states. This is somewhat surprising given the proximity of His βD4 to the aromatic ring current effects of phosphotyrosine.

Singer & Forman-Kay also investigated the pK_a of phosphotyrosine in the free

and bound states by observing the ^{31}P chemical shift of the phosphate group. A pK_a of 6.1 ± 0.1 was obtained for the free peptide, which is close to that of free phosphotyrosine, and that of other phosphotyrosyl peptides. However, for the SH2–peptide complex, ^{31}P chemical shifts were constant between pH 7.5 and 5.3; dependence on pH was only observed below this value, resulting in a pK_a of 4.0 ± 0.2. Clearly the PO_4^{2-} form is more energetically favourable during interaction with the SH2 domain. This can be rationalised from the atomic structure in which phosphate forms many hydrogen bonds within the positively-charged pocket.

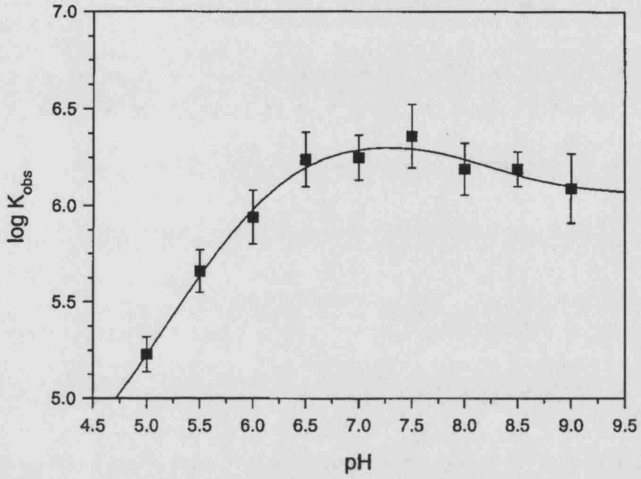


Figure 4.16 Dependence of the association constant between *v*-Src SH2 and PQpYEEIPI on solution pH. Figure reprinted with permission from *Biochemistry* **1998**, 37, 44, 15400-15407. © 1998 American Chemical Society.

The following year, Bradshaw & Waksman used isothermal titration calorimetry to characterise the pH-dependence of binding for the *v*-Src SH2–PQpYEEIPI complex[193]. They observed a significant increase in affinity (14-fold rise in K_{obs}) when raising the pH from 5.0 to 7.5, and a slight affinity decrease above this pH (Figure 4.16). The authors concluded that two ionisation events were affecting binding affinity. The first, at lower pH, shows strong linkage to K_{obs}, whilst the second, at higher pH, is clearly less influential. The data was fitted to a double-proton linkage model, obtaining pK_a values for the free and bound states of both ionisation events. For the group that ionises at low pH, free and bound pK_a values were calculated to be 6.2 ± 0.2 and 4.4 ± 0.7, respectively – a significant decrease in pK_a on binding. On the basis of Singer & Forman-Kay's report of the PLCγ C-SH2 domain, Bradshaw & Waksman assigned these pK_a values to the phosphate group of the peptide. In support of this, the estimated value of 6.2 ± 0.2 for the free state agreed well with a direct measurement of the phosphotyrosyl peptide pK_a (6.1 ± 0.1). Results from a second series of experiments, which used ITC to measure binding thermodynamics over the pH range 4.5-8.0 in several buffers of different ionisation enthalpies,

indicated free and bound pK_a values for the low-pH transition to be 6.1 ± 0.1 and 4.4 ± 0.3, respectively. These data agreed well with the first analysis, yielding a more precise value for the bound phosphotyrosyl pK_a. By comparing the PLCγ_1 C-SH2 and *v*-Src SH2 binding sites, the pK_a of His βD4 was predicted to be below 4.0, and thus no other group was likely to cause the proton linkage observed below pH 7.5.

Bradshaw & Waksman were less confident about the high-pH ionisation, with free and bound pK_a values calculated to be 8.2 ± 0.7 and 8.5 ± 0.7, respectively. It is actually possible to draw a straight line through the data from pH 6.5-9.0 and still be within experimental error, and therefore affinity may not linked to a particular ionisation event occurring at high pH. The estimated pK_a values in the range of 8-9 suggest involvement of a cysteine side chain. *v*-Src SH2 contains three cysteines, one of which (C42 or Cys βC3) is found in the phosphotyrosine pocket, and may form a weak hydrogen bond with the phosphorylated ligand. The presence of a thiolate anion could destabilise binding of the negatively charged phosphate, and lead to the slight decrease in K_{obs} observed at higher pH values. One would expect the pK_a of this group to be slightly lower than normal, due to neighbouring positive charges, but that on binding, through formation of a hydrogen bond and burial by peptide, its pK_a would increase. This is a rational interpretation of the predicted pK_a changes on binding. However, the relatively large error margins on fitted parameters prevented confident assignment of this side chain to this arguably insignificant transition. Indeed, one might expect the loss of a hydrogen bond, albeit weak, and introduction of a negative charge might cause as much a loss of affinity as protonation of the phosphate ion to the -1 form, observed at lower pH.

A cysteine is conserved at this position in Src SH2, yet absent from the closely homologous SH2 domains of the wider Src family, which feature a conservative C→S substitution[99]. Outside of this group, Cys βC3 is only found in the SH2 domains of SAP and EAT2, which are closely related to one another[2]. A C42S mutant of *v*-Src SH2 displayed a ~4-fold increase in peptide affinity over wild type[106], presumably because hydroxyl forms a more energetically-favourable hydrogen bond with phosphate than sulphydryl. Surprisingly, a C42A mutation that abolishes this hydrogen bond conferred a 7-fold improvement in affinity. This may be explained by either formation of a novel hydrophobic interaction, or alteration of desolvation effects (Dr Mark Williams, personal communication). Thus C42 appears to have an important role in modulating the affinity or function that is particular to Src, and possibly SAP and EAT2.

A separate report that described a weak phosphatase activity associated with the Src SH2 domain suggested that C42 may, at least *in vitro*, act as a nucleophile[298].

The SH2 domains of Grb2 and GAP did not share this catalytic ability, which was inhibited by orthovanadate (a classical protein tyrosine phosphatase (PTP) inhibitor) and a range of phosphotyrosyl peptides. In support of this hypothesis, it is interesting to note that PTP-mediated hydrolysis of phosphotyrosine is reliant on a conserved cysteine side chain[299]. However, a C→A mutation in the Src SH2 domain was associated with an 85 % loss of phosphatase activity, indicating that while C42 is important, replacement with alanine does not completely abolish the slow hydrolytic function. Moreover, kinetic profiles appeared to be independent of pH across the range 5.9 to 9.4, suggesting that acid-base catalysis may not be operating. Finally, it is possible that while the reaction occurs *in vitro* it may not have any relevance to *in vivo* signal transduction. It would be interesting to observe what effect a C42A Src SH2 mutant would have on, for example, hydrolysis of the autoinhibitory peptide (pY527) that would favour self-activation of the kinase domain.

In order to address these matters more fully, with particular attention to the ionisation states of His βD4, Cys βC3, and phosphate within the *v*-Src SH2 system, we have attempted to characterise the pK_a of ionisable groups within the apo and peptide-bound forms of *v*-Src SH2 using NMR chemical shift as a probe for ionisation events.

4.3.2 pH dependence of chemical shift

Ionisation events can be monitored and quantified by measuring the pH-dependence of chemical shifts. Most side chain labile protons are in very fast exchange on the NMR timescale, thus as the equilibrium ionisation state of a population of structures changes, the chemical shifts of surrounding nuclei become increasingly perturbed. Since the effect is dependent on distance, titration data from one nucleus can report the ionisation of multiple side chains. In order to obtain the most reliable data, it is preferable to monitor changes for the nucleus directly involved in, or one to two bonds separated from, the ionisation event. There are many examples in the literature of ^{13}C, ^{15}N, and ^{31}P-NMR spectroscopy being used to measure pK_a within biological structures[256, 300, 301, 302].

Resonances that show pH dependence can be fitted to a sigmoidal equation (4.12) that permits calculation of pK_a. δ_{xh} and δ_x are the chemical shifts of the protonated and deprotonated states, respectively. In more complicated situations where a nucleus experiences two unrelated transitions, shift changes can be fitted to a modified equation (4.13) in which δ_0 and δ_2 represent the chemical shifts of the protonated form of the first titratable group and the deprotonated form of the second, respectively, δ_1 is the point of inflection, and pK_{a1} and pK_{a2} are the fitted

pK_a values.

$$\delta = \frac{\delta_{xh} + \delta_x 10^{(\mathrm{pH}-\mathrm{p}K_a)}}{1 + 10^{(\mathrm{pH}-\mathrm{p}K_a)}} \tag{4.12}$$

$$\delta = \frac{\delta_0 + \delta_1 10^{(\mathrm{pH}-\mathrm{p}K_{a1})} + \delta_2 10^{(2\cdot\mathrm{pH}-\mathrm{p}K_{a1}-\mathrm{p}K_{a2})}}{1 + \delta_1 10^{(\mathrm{pH}-\mathrm{p}K_{a1})} + \delta_2 10^{(2\cdot\mathrm{pH}-\mathrm{p}K_{a1}-\mathrm{p}K_{a2})}} \tag{4.13}$$

4.3.3 Experimental approach

In order to probe the ionisation behaviour of each titratable group within *v*-Src SH2, we decided to isotopically label every carbon and nitrogen, and observe chemical shifts for each atom using 2D $^1H^{15}N$-HSQC and $^1H^{13}C$-HSQC experiments. The *v*-Src SH2 sequence contains 41 titratable groups (see Table 4.3). In order to make useful comment on the work of Bradshaw & Waksman, we employed the same pH range (5–9) and similar buffer conditions[‡]. This involved use of a triple-buffer system designed to preserve the ionic strength of the sample across the pH range. $^{13}C/^{15}N$-labelled protein was concentrated to 0.45 mM and dialysed extensively against the triple-buffer at pH 7.00. Four 0.5 mL NMR samples were then prepared, with 50 μL (10% [v/v]) D_2O added to each for signal locking. Lyophilised peptide (Ac-PQpYEEIPI-NH_2) was added to two samples to a final concentration of 1.2 mM (2.4-fold excess over protein). The pH of all samples was checked and adjusted to 7.00 as necessary before experiments began.

Titratable group	Typical pK_a	Residue number
Asp (4)	4.5	47,49,65,92
Glu (7)	4.6	3,4,14,16,23,33,35
His (3)	6.3	58,90,96
Cys (3)	9.3	42,95,102
Tyr (6)	9.7	6,41,59,70,86,87
Lys (8)	10.4	9,38,52,57,60,63,89,106
Arg (8)	~12	12,13,17,26,32,62,74,97

Table 4.3 List of titratable groups within the *v*-Src SH2 domain, excluding N- and C-termini, which have pK_a values of 8.0 and 3.1. Source: Tanford[303] and Stryer[293].

$^1H^{15}N$-HSQC and constant-time $^1H^{13}C$-HSQC (aliphatic and aromatic) spectra were collected at 0.5 pH unit intervals between pH 5.0 and 9.0 as described in Section 2.3.2. The pH of the sample was adjusted by addition of small amounts of 0.1 M NaOH or HCl as appropriate. Two *v*-Src SH2 samples, one with and the other

[‡]25 mM sodium acetate, 25 mM MES, 50 mM TRIS, 50 mM NaCl and 1 mM DTT. Bradshaw & Waksman used β-mercaptoethanol as a reducing agent instead of DTT.

without PQpYEEIPI, were used to collect data above pH 7.00 and an equivalent pair of samples were used to obtain data for pH values below 7.00. FID data were processed as described in Section 2.3.3 and spectra were analysed in *ANSIG* v3.3 [226]. Spectra describing the apo domain at pH 6.00 were very similar to those obtained in MES buffer during structure determination.

Data analysis

Raw chemical shift data was fitted by *ORIGIN* (*OriginLabs*) to either Equations 4.12 or 4.13 depending on whether a single or double sigmoidal transition was apparent. Two problems were encountered whilst analysing data. Firstly, the presence of five protonated small molecule species (acetate, MES, TRIS, DTT and H_2O) caused difficulties in processing and interpretation of the aliphatic constant time $^{1}H^{13}C$-HSQC. When the SH2 domain resonances were correctly phased, the intense buffer peaks were slightly dephased, resulting in general spectral distortions and occlusion of some protein resonances. As buffer chemical shifts were pH-dependent, the effects of this were variable at different data points. Complications arising from buffer peaks could have been avoided by using deuterated (i.e. $^{2}H^{12}C$) buffers; this is highly recommended for future studies of this type.

A second complicating issue was the adverse rates of solvent exchange for backbone amide protons experienced at pH values above 7. For instance, approximately 30 % amide resonances were broadened beyond detection by pH 9. As a result, complete chemical shift data sets could not be obtained for every amide. However, the fact that certain amides are protected against solvent exchange provides useful information regarding burial or the formation of hydrogen bonds, as discussed below in Section 4.4.

4.3.4 Results from the ^{1}H-^{15}N HSQC spectra

Since backbone amides are spread evenly throughout the protein they are useful reporters of protein-ligand interactions, or in this case, ionisation events. However, their exquisite sensitivity towards the electromagnetic environment can lead to problems in the quantitative interpretation of pH-dependent shift changes because of the greater degree of interference between ionisation events. Despite this disadvantage, it is useful to consider the effect of pH on each amide in the free and bound states in order to visualise local and global changes. Changes in ^{1}H and ^{15}N shifts between the pH 5.0 and 9.0 (or highest pH at which the resonance is still resolvable) were measured and combined, as described in Section 4.1.1.1. Values were plotted against residue sequence and the shift changes used to colour the apo *v*-Src SH2 structure (Figure 4.17). The largest perturbations are located in the vicinity of

H90/H96, suggesting that ionisation of these, and possibly neighbouring titratable groups, leads to the most significant changes in electromagnetic environment. Titration of pH also led to chemical shift changes for particular amides in and around the phosphotyrosine binding site.

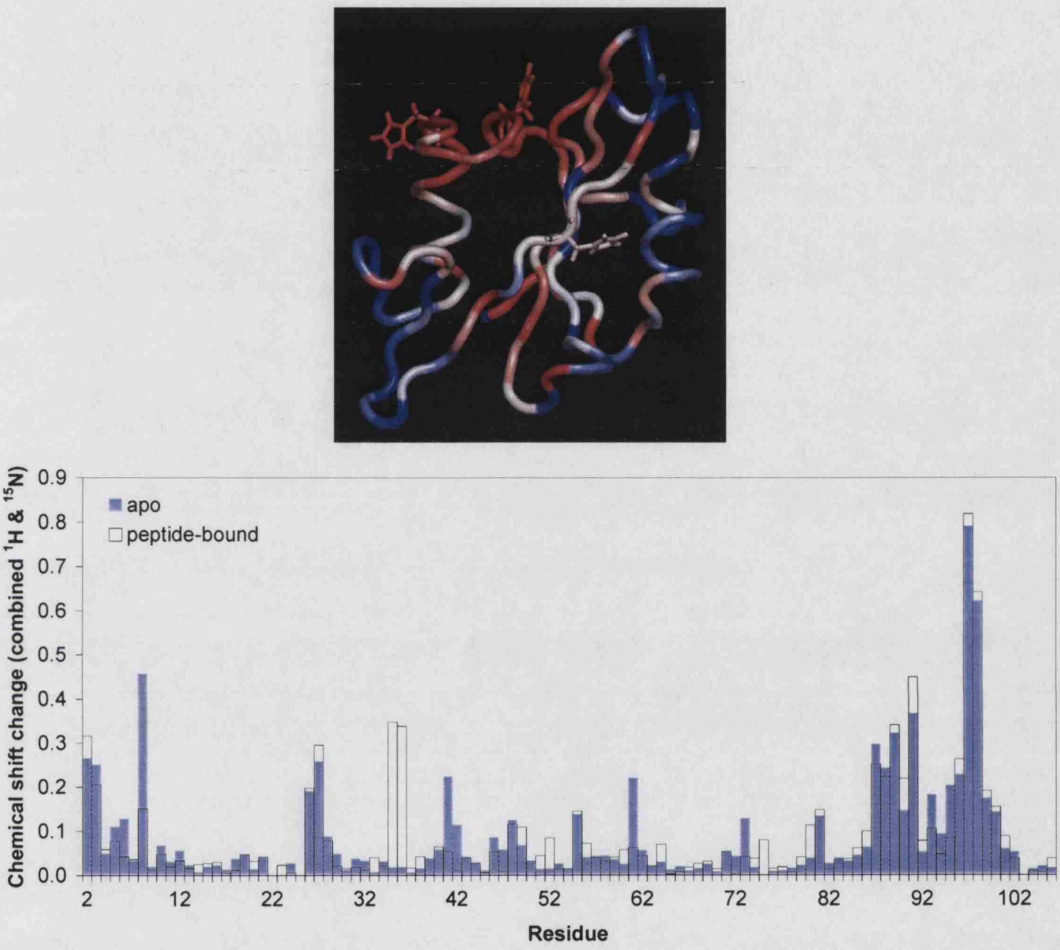


Figure 4.17 pH-induced chemical shift changes in $^{1}H^{15}N$-HSQC resonances. Changes are expressed as the difference in combined ^{1}H and ^{15}N coordinates between pH 5 and 9 (or highest available) spectra, for the apo and peptide-bound states. *Top* | Total observed shift changes for the apo domain; the backbone is coloured in a blue–white–red gradient, proportional to the extent of shift difference. Histidine side chains are shown as sticks. *Bottom* | Bar graph displaying amide shift changes for the apo (filled blue bars) and peptide-bound (unfilled bars/black boundary) states.

Interestingly, there are a few regions in which the presence or absence of peptide is found to affect local sensitivity to pH, suggesting that some pK_a values change on binding. The amides of Y41-C42, I61, and I71-S73, located in three adjacent strands of the central β-sheet, are a good example of this phenomenon. In the apo state, these amide resonances display a significant and similar perturbation between pH 7.5

and 9, however peptide binding abolishes this effect. A difference in the pK_a of C42 thiol may explain this finding. Protonation events at this site will be experienced by neighbouring amides, and in this case the effect appears to be propagated across the β-sheet. Burial of the thiol group by the peptide and formation of a hydrogen bond to phosphate, would both cause an increase in its pK_a, and account for the absence of pH-sensitivity in the bound state. A more reliable measurement of C42 pK_a is required in order to substantiate this hypothesis.

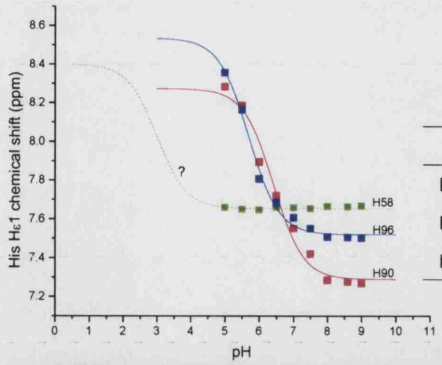
The other obvious case in which peptide binding reduces the magnitude of pH-related amide chemical shift change is for G8. In the apo domain, this site reports an ionisation occurring at high pH that, for reasons of structural proximity, is highly likely to represent dissociation of Y6 hydroxyl. The environment around G8 does not appear to alter between the apo and peptide-bound states, so it is not immediately apparent why equivalent chemical shift changes do not occur in spectra of the complex. The interaction with peptide must in some way influence either G8 or Y6 such that the ionisation is not sensed, or perhaps increases the pK_a beyond the range of detection in this investigation.

An opposite effect is found for E35/T36 where peptide binding appears to introduce a new ionisable group (with p$K_a \leq 5.5$) in the vicinity of these residues. This can be readily assigned to the ionisation of phosphate, which is very close to these residues in the bound state.

4.3.5 Histidine pK_a values

The imidazole pK_a for free histidine is 6.0, however removal of the zwitterion charges by N-acetylation and C-methylation, or formation of peptide bonds raises the value to 6.3-6.4.[304] Of the three histidines within the domain, two – H90 and H96 – are close in sequence and space (inter-ring distance approximately 11 Å) and are exposed to solvent (see Figure 4.17). The third residue (H58 or HisβD4) forms part of the phosphotyrosyl binding site and is partially buried. Histidine $H_{\epsilon 1}$ chemical shift data were used to obtain best-fit pK_a values describing the free and bound states (see Figure 4.18).

Within the apo *v*-Src SH2 structure ensemble, the side chain of H90 is completely solvent-exposed. This is supported by its observed pK_a value, which matches exactly that observed for free histidine residues. The free/bound pK_a values of H96 are slightly more acidic, a finding that can be rationalised by the presence of two flanking, positively charged arginines (R26 & R97). Peptide binding appears to partially relieve this modulatory effect, perhaps through loss of a salt-bridge between K57 and D47, leading to altered hydrogen bonding between D47 and R26 that results in movement of the guanidinium moiety away from H96. The increased



	Apo pK_a	Bound pK_a	ΔpK_a
H58	< 4.0	< 4.0	?
H90	6.38 ± 0.17	6.62 ± 0.11	+0.2
H96	5.70 ± 0.12	6.07 ± 0.07	+0.4

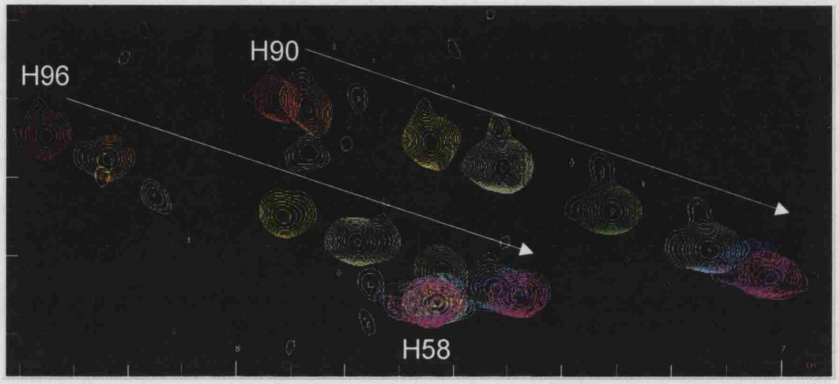


Figure 4.18 Analysis of histidine pK_a values. *Bottom* | Overlay of a selected region of aromatic $^1H^{13}C$-HSQC spectra recorded at pH intervals using the peptide-bound domain. Each histidine displays a single ϵ_1CH resonance, labelled in the pH 5 spectrum (red contours). Increasing pH towards 9 leads to a reduction in ^1H chemical shift for H90 and H96, indicated by the white arrows. Conversely, the chemical shift of H58 is insensitive to pH across this range. *Top* | Proton chemical shift coordinates for the ϵ_1CH resonances were plotted (*left*) and used to obtain appropriate pK_a values describing each transition for the apo and bound states (*right*). Reliable fits could not be obtained for H58, suggesting that its pK_a must be below 4.0, as shown by the hypothesised dashed titration curve.

solvent-accessibility of H96 during this proposed sequence of events would also serve to increase its pK_a above that of the apo form.

His βD4 (H58) presents an entirely different situation since chemical shifts for this residue, and other neighbouring nuclei, were not sensitive to pH across the sampled range in either the apo or bound states. Ionisation events appear to manifest themselves across a range of at least pH two units[§], therefore the pK_a value of H58 is most probably below 4.0. This important finding corroborates with data describing the PLCγ_1 C-SH2 domain and strongly suggests that in both the apo and bound structures of *v*-Src SH2, His βD4 and Arg βB5 are connected by a hydrogen bond.

As discussed above, Bradshaw & Waksman attributed the significant dependence of peptide affinity on pH between 5 and 7.5 to the secondary ionisation of phospho-

[§]This opinion was arrived at through empirical analysis of these data

tyrosine, rather than (de)protonation of His βD4. They assumed that the pK_a of His βD4 would be similar to that observed in PLCγ_1; these results prove this supposition to be valid.

4.3.6 Cysteine pK_a values

The pK_a of a solvent-exposed cysteine thiol group is often found to be in the range 9.1-9.5. Glutathione, a small pseudopeptidic molecule that contains cysteine has a pK_a of 9.2. Due to the useful chemical properties of the thiolate anion, the pK_a of particular cysteine side chains have been found to be lowered by several pH units (e.g. glutaredoxin[305], DsbA[306], and protein tyrosine phosphatases[307]). *v*-Src SH2 contains three cysteine side chains at positions 42, 95 and 102, corresponding to Cys βC3, Cys BG5 and Cys βG4, respectively. The βCH resonances of C42 showed dependence on pH, which were fitted to a pK_a value of 8.56 $\pm$ 0.42 in the apo domain (Figure 4.19). Bradshaw & Waksman suspected that the pK_a of C42 would be below that of solvent-exposed cysteine since it is close to the guanidinium group of R32. This measurement supports their hypothesis. In the peptide-bound state, the βCH resonances are distorted by overlap with buffer peaks, however the αCH peak could be fitted to the pK_a equation, generating a value of 8.80 $\pm$ 0.17 (Figure 4.19). Although the calculated pK_a values for the apo and bound states are within error of each other, there are good structural reasons to support a slight increase in Cys βC3 pK_a on binding through burial by the peptide and formation of a hydrogen bond with phosphate.

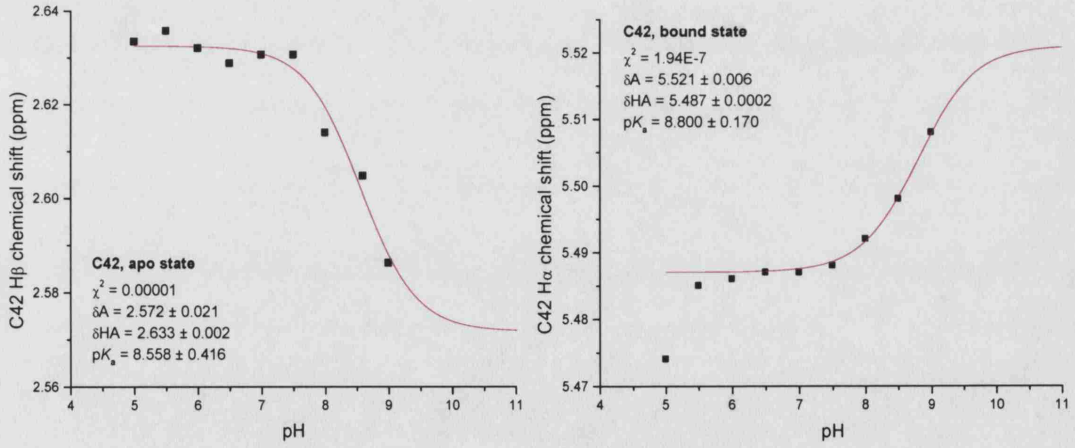


Figure 4.19 Calculation of C42 pK_a values for the apo (*left*) and bound state (*right*). The latter pK_a was fitted using Equation 4.12 (single transition) as it produced a more accurate value than if the low-pH transition was included.

Bradshaw & Waksman detected a slight decrease in peptide affinity above pH 8 (see Figure 4.16 on page 175), suggesting linkage to an ionisation event that has

pK_a values of 8.2 ± 0.7 and 8.5 ± 0.7 describing the apo and ligand-bound states, respectively. The NMR approach employed in this study generated very similar pK_a values for C42 in the free and bound states, which strongly suggests that loss of a proton from the thiol at relatively high pH (8–9) is responsible for the slight reduction in peptide binding affinity. The fact that a C42S mutant has been reported to display slightly higher affinity than wild type supports this conclusion. It would be interesting to study the relationship between pH and K_{obs} with the C42S mutant. If the C42 ionisation hypothesis is correct, affinity of the domain for peptide should be independent of pH over the range 7.5–9.

Unfortunately, an intense Tris buffer resonance was degenerate with the βCH resonances of C95 and C102, between pH 8-9, which prevented detection of reliable chemical shift changes for these nuclei. The αCH resonance of C102 displayed no pH-dependent perturbations, suggesting that its pK_a is fairly high ($\geq$10). The observation that C102 is buried within a hydrophobic environment created by L18, L19 and L30 supports the assignment of a high pK_a. Similarly, in the bound state, resonances for C102 appear to be insensitive of pH; the increased pK_a is sustained during interaction with peptide.

Data for the apo C95 αCH resonance was fitted to a pK_a value of 8.5 ± 0.1, which is slightly lower than expected for solvent-exposed cysteine. This deviation can be explained by the proximity of two positively charged groups (R26 & K57). Binding of peptide caused the αCH resonance to overlap with others, preventing further unambiguous detection. Other surrounding nuclei did not display pH-dependent chemical shifts and therefore a pK_a value could not be obtained for C95 in the bound state.

4.3.7 Phosphate pK_a values

The pK_a value describing the second phosphate ionisation ($HPO_4^- \rightleftharpoons PO_4^{2-} + H^+$) was obtained by measuring the ^{31}P chemical shift across the pH range 3.5 to 9.0¶. Raw data (not shown) was collected for peptide alone (2 mM) and for the SH2–peptide complex at a stoichiometry of 1:2. The resultant free and bound state pK_a values were 5.99 ± 0.02 and 4.90 ± 0.18, respectively, indicating a significant decrease in phosphate pK_a on binding. These values agree with previous phosphotyrosine pK_a measurements for *v*-Src SH2 by ITC[193] and PLCγ SH2 domain complexes by NMR[256].

¶Chemical shift data was collected and fitted to pK_a equations by Dr A. Ababou using SH2 domain prepared by the author

4.3.8 Summary of pK_a investigation data

This investigation into the ionisation states of particular side chains in the *v*-Src SH2 SH2 domain has provided information on several key points of interest. Firstly, the unusually low pK_a value determined for His βD4 (H58) validates previous structural information for the apo and peptide-bound states of *v*-Src SH2, which predict a hydrogen bond between this side chain and Arg βB5 (R32). This probably functions to restrain Arg βB5 conformational flexibility such that it is primed to interact with phosphotyrosine. Since both residues are highly conserved throughout the SH2 domain family, it would seem likely that the hydrogen bond connecting them is equally conserved. In a recent study, Campbell & Jackson demonstrated the structure of the phosphotyrosine binding site is particularly well-conserved. However, to date the His βD4–Arg βB5 interaction has only been reported explicitly for the C-terminal SH2 domain of PLCγ_1 [99, 256].

Despite the strong inference from bioinformatic data, some *in vitro* and structural studies suggest that this hydrogen bond may not be conserved nor important to phosphotyrosine recognition. Mutation of His βD4 to alanine resulted in a two-fold decrease in affinity compared to wild type[106]. In comparison to the effect of other substitutions (e.g. R12A, R32A, K57A), His βD4 does not appear to be crucial in determining high affinity for peptide. Moreover, inspection of SH2 domain apo and peptide-bound structures within the Protein Data Bank revealed that Arg βB5 and His βD4 are not always oriented in a manner that is conducive to mutual hydrogen bonding. Of the 18 apo structures available, 6 predict this interaction, whilst a further 10 did not meet the criteria for hydrogen bond geometries[‖]. In the latter group of structures, appropriate rotation and/or translation of His βD4 and Arg βB5 side chains would bring them within hydrogen bonding distance. If it is believed that these two side chains interact in the apo state, then nearly two thirds of the determined structures would have to be described as 'incorrect', for this site at least.

A similar analysis of the 18 unique SH2 domains that feature a ligand bound to the pY pocket revealed 10 cases in which the His βD4–Arg βB5 hydrogen bond is possible, and a further 3 domains which could form this bond if one or both side chains were moved. Five of the SH2 complex structures determined to date feature a substitution at His βD4. Interestingly, one of these domains (Cbl) indicates a hydrogen bond between a neighbouring asparagine (i.e. βD3) and Arg βB5, suggesting that the hydrogen bond is important and conserved. Thus the hydrogen bonding that serves to order the side chain of Arg βB5 appears to be more prevalent within structures of SH2-complexes than of apo domains. The fact that some domains have a greater distance between His βD4 and Arg βB5 could be explained by errors within

[‖]Histidine substitutions were observed for the final two domains

the structure determination process. For example, some of the structures specified the doubly protonated form of histidine, which could not hydrogen bond with arginine in structure calculations. Alternatively, it is quite possible that the hydrogen bond may not be as important in some domains as in others. Measurement of the relevant pK_a values for HisβD4 in these domains should help to confirm or deny a specific interaction with ArgβB5. It is useful to know the structure and protonation states of binding sites for structure-based drug design and molecular docking. The knowledge that HisβD4 is already hydrogen bonded to another residue would prevent time being wasted in designing inhibitors targeting this side chain.

This pK_a investigation has also lent strong support to Bradshaw & Waksman's prediction regarding slight decreases in affinity above pH 8. The thiol pK_a of C42 is lower than normal in the apo state, consistent with its reactivity exploited in cysteine-targeted inhibitors, however this is found to increase during interaction with peptide, which indicates a preference for a hydrogen bond between C42 and phosphate. Whilst the contribution of C42 towards *in vitro* binding has been clarified by these observations, the question of why C42 is substituted by serine in other Src family SH2 domains remains unanswered. The study by Bradshaw & Waksman also predicted that the phosphate pK_a decreases from 6.1 $\pm$ 0.1 to 4.4 $\pm$ 0.3 on association with peptide. Once again, direct measurement of the phosphate pK_a in this research has confirmed their prediction.

4.4 Effect of peptide binding on ^{1}H-^{15}N HSQC intensities

During the analysis of $^{1}H^{15}N$-HSQC spectra in the pK_a investigation, a variable sensitivity to pH was noted for the intensity of backbone amide resonances. For example, the E35 peak became indistinguishable from spectral noise between pH 7.5 and 8.0 whilst that of L30 remained clearly visible at pH 9.0. Moreover, addition of peptide was found to alter pH-sensitivity of amide resonances, and revealed further information regarding the effects of binding. The appearance of an amide resonance in the HSQC spectrum is dependent a number of factors that affect $R_{2,obs}$ and solvent exchange rates. In the former case, exchange processes can have a significant effect on $R_{2,obs}$, which directly modulates linewidth. For example, very rapid conformational exchange is associated with narrow linewidths whilst slow-intermediate exchange regimes have a negative impact on linewidth. The solvent exchange rate, defined as the rate of chemical exchange between labile protons and solvent protons, is controlled by pH (due to catalysis by OH^-) and solvent accessibility, which is related to structure and dynamics. At low pH, amides undergo slow chemical exchange, leading to narrow linewidths at the amide and H_2O frequencies. As the exchange rate increases, the process enters the 'intermediate' timescale, causing lines

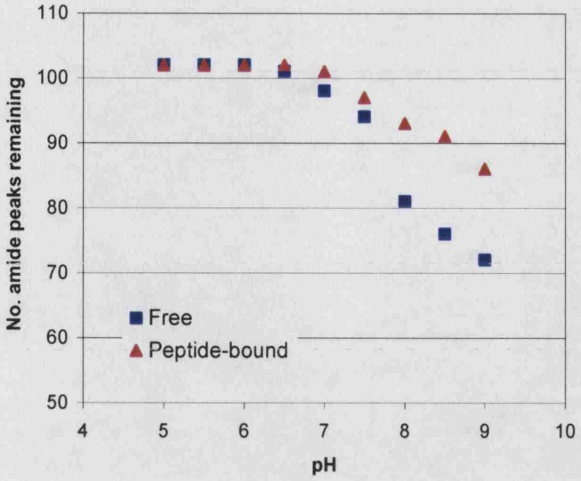
to broaden and eventually disappear into the noise. Research into solvent exchange rates of biomolecules has indicated that amides forming hydrogen bonds, or those that have low solvent accessibility, receive a varying degree of protection against exchange (see Krishna *et al.*[308] for a recent review).

The interaction between *v*-Src SH2 and PQpYEEIPI caused a change in the sensitivity of amide resonances to pH. This can be explained by invoking some degree of protection against solvent exchange or by considering the alteration in local dynamics induced by peptide binding. Additional information in the form of structure, hydrogen bonding, and relaxation rates helps to decipher the contributions of each process to the observed change in pH-sensitivity.

For the apo *v*-Src SH2 domain, 30 amide resonances that are present in the spectrum at pH 5 are absent at pH 9 (Figure 4.20). These amides are largely confined to loops and surface-exposed positions. The binding of phosphopeptide leads to a ~50 % reduction in the number of resonances that are absent from the pH 9 spectrum. Moreover, many of the changes in pH-sensitivity are located in and around the peptide binding sites suggesting that the peptide induces some form of protection against solvent exchange at these sites. Peptide binding was not observed to cause an increase in sensitivity to pH for any amides.

The persistence of individual amide resonances in the HSQC spectrum at increasing pH values permitted ranking of this protective effect. For instance, the T36 amide resonance is absent from the apo spectrum above pH 7, yet is still visible in peptide-bound spectrum at pH 9, indicating a 'degree of protection' of +2.0 pH units. Similarly, the R26 resonance disappears above pH 8 in the apo spectrum, however in the presence of peptide, it remains observable at pH 9, yielding a protective value of +1.0. Degrees of protection were derived in this way for all residues: 77 amides generated a value of zero whilst 12, 7, 5 and 1 amide(s) displayed +0.5, +1.0, +1.5, and +2.0 units, respectively. The identity and spatial position of amides with non-zero degrees of protection are shown in Figure 4.21. It is clear that several protected sites occur in and around the peptide binding site, reflecting exclusion from the solvent-exposed surface, or perhaps formation of new hydrogen bonds.

A detailed analysis of the apo (NMR) structure and peptide-bound (x-ray) structure indicated that binding of peptide is associated with the formation of several hydrogen bonds that are likely to provide additional protection against solvent exchange. For example, the amide of E35 displays a degree of protection of +1.5 as it is solvent-exposed in the apo state but hydrogen bonded to a phosphate oxygen in the bound state. The behaviour of T36 (+2.0) could be explained in an identical way. K60 is protected by the formation a hydrogen bond to a trapped water molecule, which also interacts with the pY+1 backbone carbonyl group. Peptide-binding also



pH	No. peaks visible in ^{1}H-^{15}N HSQC apo	+ PQpYEEIPI
5.0	102	102
5.5	102	102
6.0	102	102
6.5	101	102
7.0	98	101
7.5	94	97
8.0	81	93
8.5	76	91
9.0	72	86

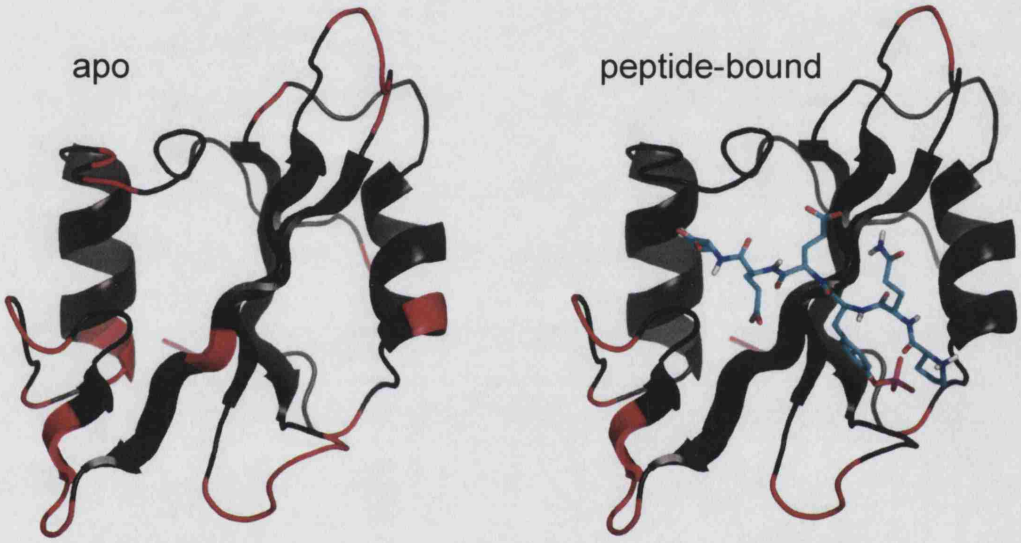


Figure 4.20 Effect of pH on $^{1}H^{15}N$-HSQC peak intensities. *Top* | Graph and associated table showing the total number of amide resonances visible in the free (blue squares) and peptide-bound (red triangles) spectra between pH 5 and 9. *Bottom* | Amide resonances absent from the pH 9 spectra are coloured red in the apo (left) and peptide-bound (right) structures.

leads to an alteration in hydrogen bonding partner of D49 carboxylate: In the apo state, this side chain interacts with the ammonium group of K52, however in the bound state D49 forms two new hydrogen bonds with the backbone amides of A51 and K52, leading to protection against solvent exchange. The glycines of the DE loop, G67 & G68, both form hydrogen bonds with the backbone carbonyl group of L64 in the peptide-bound state, yet are solvent-exposed in the free state, which accounts for their medium degrees of protection (+1.0) against solvent exchange. Finally, the peptide-induced protection of Q81 amide can be explained by noting the formation of a hydrogen bond to its own side chain $O_{\epsilon 1}$ that is absent in the free state.

Degree of protection	Colour	Residues
+0.5	Green	2,3,4,9,38,50,54,65,66,78,104,106
+1.0	Blue	26,51,52,67,68,74,91
+1.5	Yellow	13,35,60,81,93
+2.0	Red	36

Figure 4.21 Degrees of protection for amides are assigned according to the difference in pH values at which they disappear from the apo and peptide-bound spectra. The magnitude and location amides displaying non-zero values are shown in the table, which includes the key to protein colouration.

The amides of R13, R26, R74, A91 and G93 do not appear to form intramolecular hydrogen bonds in either the apo or peptide-bound structures, yet in the presence of peptide, their resonance intensities persist at higher pH. This is probably because interaction with peptide causes a change in local dynamics that reduces linewidth and permits continued detection of the resonances above the noise.

Insertion of the peptide isoleucine into the pY+3 pocket was found to alter the dynamics of the EF and BG loops. Relaxation data described above shows that compared to the population mean, the motional rates (as evidenced by R_2/R_1 ratios) of R74, A91, and G93 amides are smaller in the presence of peptide than in the absence, which may have a favourable effect resonance intensity. Similarly, hydrogen bonding between R12 side chain and phosphate probably restrains the N-terminal end of the αA helix, resulting in a change in dynamics for R13 amide.

R26 presents a more interesting situation as it is further from the binding site, yet is indirectly connected with the effect of peptide interaction. In the apo state, the carboxylate side chain of D47 interacts with both the guanidinium groups of R26 and ammonium group of K57 (Figure 4.22). Upon peptide binding, E+1 carboxylate sequesters the ammonium group of K57, permitting some realignment of R26 and D47, which remain hydrogen bonded to each other. This may have an effect on the dynamics measured at R26 amide, however the relaxation data does not show a significant perturbation associated with peptide binding. Chemical shifts for R26 differ slightly between the apo and bound states (Section 4.1.3) and the altered pK_a of H96 is thought to depend upon movement of R26 away from this imidazole group (Section 4.3.5). Thus structural changes associated with peptide binding are

implied, however results were not so conclusive regarding dynamics.

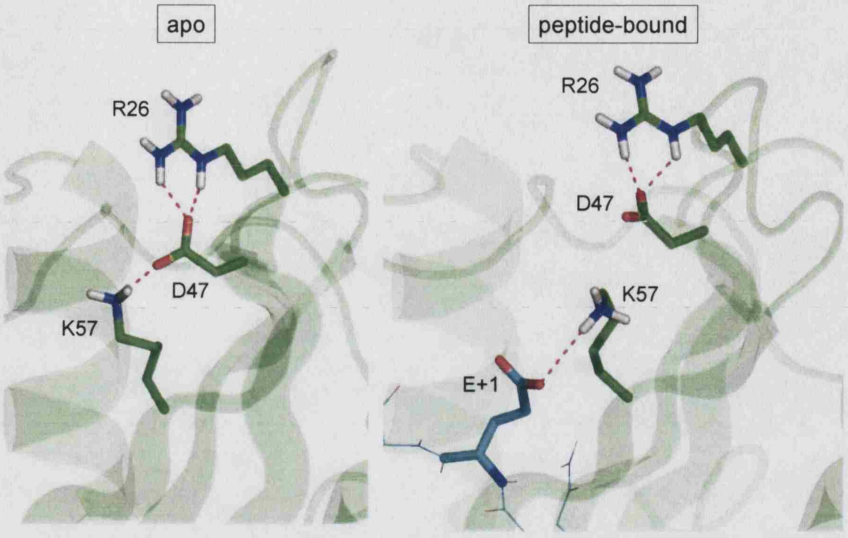


Figure 4.22 Peptide binding leads to an alteration in hydrogen bonding, with proposed effects on R26 amide relaxation rate.

Whilst this analysis of $^{1}H^{15}N$-HSQC peak intensities at various pH values has not yielded novel insights into SH2 domain binding behaviour, it has provided supportive data for existing structural information concerning the apo and peptide-bound states of *v*-Src SH2. As far as the author is aware, this approach is not generally employed in the analysis of solvent exchange rates for backbone amides. Most likely this is because it is difficult to distinguish between the effects of protection against solvent exchange and the changes in R_2 that affect signal intensity. In this case, certain changes in resonance behaviour could be readily interpreted by invoking additional hydrogen bonding patterns. However the conclusions were less obvious in cases where such interactions appear to be unlikely. H/D exchange and amide proton temperature coefficients remain more common and powerful approaches to investigating hydrogen bonding; this non-standard method could be employed to provide additional supportive evidence.

4.5 Summary of chapter

Nuclear magnetic resonance techniques have been used to investigate the nature of the apo SH2 domain and its complexes with phosphate, phosphotyrosine, and PQpYEEIPI. The chemical shift and relaxation analyses indicated that binding of phosphate and pY led to changes in the local electromagnetic environment and dynamic profile of nearby amides. Very few changes are observed for other regions of the domain. In the apo state, the pY pocket is more rigid than the pY+3 pocket. This is probably because the binding site is made up of secondary structural units

that are restricted in their motion and, as the pK_a investigation showed, there is a hydrogen bonding network that pre-orders the position of Arg βB5. The lack of flexibility supports the general observations from SH2-related drug design that the pY pocket is intolerant to changes in ligand structure – it is too specific. The higher degree of backbone flexibility associated with the pY+3 pocket presents the opposite problem – for synthetic ligands, this site is not specific enough. These matters are discussed more fully in Chapter 7.

Chapter 5

In silico fragment screening

In order to assess the utility of fragment NMR screening approaches, a small library of test compounds, including at least one SH2 domain ligand, was required. Whilst a number of phosphotyrosine mimetic compounds have been discovered (reviewed in Section 1.4) and would provide a useful pool of ligands for screening, we also aimed to explore new ground in the field of SH2 domain drug design by identifying novel fragment ligands. This chapter describes the selection of a small library of suitable fragment through application of computational methods. The procedure involved filtering a large database for fragments and using molecular docking and chemical similarity analyses to locate compounds that display physico-chemical complementarity towards the SH2 domain binding site and share homology with known ligands. These computational approaches were validated and optimised using a range of known fragment ligands for the SH2 domain. The first part of this chapter reviews the docking and chemical similarity searching methods employed. The remainder of the chapter describes how the software was used to identify potential fragment ligands through a combination of *in silico* screening and chemoinformatics.

Almost every step in the modern ligand discovery process involves the support of computational resources. The large compound libraries used in current HTS protocols require that databases to be kept and updated with a variety of information. Robotic systems are often employed during the synthesis and analysis of compounds, as well as in the screening process itself. Computers are also harnessed in the collection, processing, and interpretation of data using a range of automatic and manual methods. In particular, computers have proved indispensable to structure-based drug design whereby macromolecular structures complexed with different ligands are obtained and analysed. Related to this approach, a variety of computer-driven methods have been developed that attempt to predict which compounds will interact with the target of interest. This is part of a concerted attempt to reduce the size of the library to be screened *in vitro*. Prediction methods fall into two cate-

gories depending on whether they utilise structural information from the receptor or from known ligands only. The former group attempts to determine whether mutual physico-chemical complementarity exists between the receptor binding site and the compound in question. In some cases (docking, or structure-based screening), the optimal interaction between a given compound structure and the binding site is predicted. An alternative approach is to interpret the binding site as a three-dimensional array of generalised interaction sites, and used to search for ligands that match the required shape and distribution of contacts.

The second class of computational approaches attempts to identify novel ligands by searching for compounds that share structural or chemical features with known ligands. An example of this is 'pharmacophore searching'. A pharmacophore is the minimum chemical functionality a molecule has to contain in order to exhibit activity against the target of interest. This can be determined through structural comparison of known ligands and then used as a search query in compound databases. Alternatively, chemical similarity searching identifies compounds that are similar to known ligands by using quantitative molecular comparisons based on chemistry and structure.

5.1 Molecular docking and FlexX

The goal of docking software is to predict the structure and energetics of biomolecular interactions. This is a very difficult computational problem; a rigorous solution would require realistic sampling of every possible conformation for the receptor, ligand and solvent atoms. At current levels of algorithmic sophistication and computer technology, this is not yet feasible. In order to simplify the task, a number of assumptions are made regarding molecular flexibility, acceptable interaction geometry and energetic scoring. Most docking programs permit (quantised) degrees of freedom in the ligand but do not allow receptor flexibility. More recent algorithms now also specify limited flexibility in receptor structures. For instance, GOLD[309] permits rotation of protein Ser, Thr and Tyr hydroxyls, and Lys ammonium groups and FLEXE[310] incorporates alternative side-chain conformations and minor loop movements. *Celltech* use FLEXX[311], to perform compound docking. This is a fast method of predicting the interaction between small molecules and a designated macromolecular active site. The following section will serve as a description of the program concepts, operation and weaknesses.

5.1.1 Ligand conformational flexibility

FLEXX performs molecular docking whilst exploring the rotameric states available to the ligand. The conformational flexibility search engine is derived from the program MIMUMBA, which quantifies multiple low-energy conformations for small, organic molecules[312]. The bond lengths and angles in the supplied structure are not varied, however a set of preferred torsion angles are assigned to each acyclic single bond by matching the bond to equivalent fragments in the Cambridge Structural Database (*CSD*[313]) and analysing the distribution of observed angles. Commonly-occurring angles, up to 12 per bond, are selected for generation of ligand conformations. Ring puckering alternatives are computed by the program SCA[314].

5.1.2 Identifying protein-ligand interactions

Empirical intermolecular interactions are determined by fairly restrictive geometric rules, which are used by FLEXX to place ligands in the active site. In particular, FLEXX considers pairwise interactions between hydrogen bond acceptors and donors, metal acceptors and metals, and also interactions between aromatic ring centres and aromatic ring atoms, methyl groups and amides. An appropriate *interaction surface* is calculated for each *interaction centre* (Figure 5.1). Pairwise interaction between two groups are accepted if their types are compatible, and if the interaction centre of one lies on, or close to, the interaction surface of the other, and vice versa.

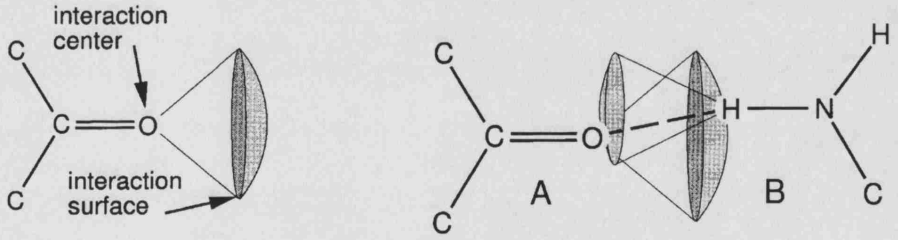


Figure 5.1 *Left* | FLEXX identifies *interaction centres*, and calculates an appropriate *interaction surface* for each. *Right* | An interaction is scored if the pair of centres (A and B) are positioned on, or close to, the associated interaction surfaces. These figures are reprinted from Rarey *et al.*[311], with permission from Elsevier ©.

5.1.3 Estimation the free energy of binding

Quantification of docking solutions in terms of free energy employs a scoring method adapted from that of Böhm[315], in which ΔG_b° is estimated using several interaction terms, as described by Equation 5.1.

$$\Delta G = \Delta G_0 + \Delta G_{\text{rot}} \cdot N_{\text{rot}} + \Delta G_{\text{hb}} \sum \alpha + \Delta G_{\text{io}} \sum \alpha + \Delta G_{\text{aro}} \sum \alpha + \Delta G_{\text{lipo}} \sum \alpha \quad (5.1)$$

The terms ΔG_{hb}, ΔG_{io}, ΔG_{aro} and ΔG_{lipo} are adjustable parameters that reflect the free energy available for optimal hydrogen bonding, ionic, aromatic, and lipophilic interactions, respectively. Deviations from ideal distances and angles are penalised using the scaling function, α. N_{rot} corresponds to the number of rotatable bonds immobilised in the complex.

5.1.4 Overall docking strategy

The FLEXX docking algorithm consists of three main stages:

1. *Base selection*: A 'base-fragment' is chosen interactively.
2. *Base placement*: The base-fragment is inserted into the active site n times, specified by the user.
3. *Complex construction*: The ligand is reassembled incrementally.

The base-fragment is a selected subset of ligand atoms that are inserted into the active site. It usually corresponds to a group that contains multiple potential interaction centres, such as a phosphate or nucleoside moiety. The remainder of the ligand is split into sub-fragments at each rotatable, acyclic single bond. For very small molecules, such as alanine, the entire compound is used as the base-fragment. Selection of this group can be performed manually, although for more extensive virtual screening exercises, automatic selection is adequate. Empirically, it is found that base selection strongly determines the success of FLEXX docking calculations – it is important that the base-fragment contains a sufficient number of interaction centres. The size and flexibility of the initial base-fragment also determines the probability of successfully predicting the binding conformation[311].

A pattern recognition algorithm is used to determine optimal placement of the base-fragment in the active site by maximising the number of pairwise interactions that can occur simultaneously. The base-fragment is treated as a rigid molecule during this stage. Multiple fragment placements are scored and ranked using the free energy estimation of Böhm[315] described above. Each accepted (and unique) base-fragment placement enters the final stage of FLEXX calculations – complex construction – in which the remaining sub-fragments are added in a step-wise fashion, using a tree-search method. The base-fragment is extended by one sub-fragment, in all possible conformations, except those that lead to significant atomic overlap. The algorithm then searches for new interactions between the protein and ligand, which

are optimised in terms of geometry. If this leads to unacceptable atomic clashes in other parts of the complex, the solution is rejected, and other possibilities are explored. After addition of each sub-fragment, the energetically best 500 placements are identified, and clustering analysis is used to group similar solutions. The highest ranking structure from each group is used in the next iteration of ligand re-assembly.

5.1.5 Preparation of input structures

The ligand physico-chemical features are defined though appropriate addition of hydrogens, assignment of formal charges, and energy-minimisation using the TRIPOS[316] force field. The receptor structure file is processed in a similar manner, except that it is not usually energy-minimised. The target docking site is defined either by manual selection of residues, or alternatively, where a co-crystal structure is available, through automatic identification of protein atoms within a set distance (usually 6.5–8 Å) of the ligand. FLEXX also obtains information about the solvent accessibility of receptor atoms through determination of the Connolly[317] surface. FLEXX is able to incorporate rigid water molecules in docking calculations, if they are known to be crucial for optimal ligand binding interactions.

5.1.6 Method weaknesses

Each step of the FLEXX docking process requires simplifications that are necessary to reduce the computational complexity of the problem. These include the models of ligand flexibility, protein-ligand interaction geometry, free-energy scoring, and the incremental construction algorithm. As well as ignoring the issue of protein dynamics, FLEXX does not predict changes in the protonation or tautomerisation states that may occur upon formation of the ligand-protein complex.

5.2 Chemical similarity searching and *DiverseSolutions*

The advent of combinatorial chemistry and ensuing creation of large compound libraries for HTS generated a problem for the pharmaceutical industry: with so many structures to choose from, which compounds should be acquired and what is the most efficient way of screening them against a given target? Debate surrounding this important question has been dominated by theories of chemical diversity and 'drug-' or 'lead-likeness'. The former concept explores the similarity and dissimilarity between chemical structures in order to maximise library diversity, minimise redundancy, and to identify and fill diversity 'voids'. Practical considerations, such

as synthetic accessibility and cost, and a reluctance to screen non-drug-like compounds all tend to reduce the diversity of compounds selected for screening.

The need for an objective assessment of chemical diversity is being addressed through application of computational methods that form part of the incipient field of chemoinformatics*. Any concept of chemical similarity is reliant on some quantitative comparison of structures located at positions in multi-dimensional 'chemistry space', the axes of which are termed 'metrics'. A variety of molecular descriptors can be used as metrics, such as charge and hydrogen bonding potential, however these need to be validated on an individual basis, as described below. Initial attempts to classify compounds involved assignment of binary 'molecular fingerprints', which could be compared using the Tanimoto[318] similarity index[319]. Essentially, each 1 or 0 'bit' represents the answer to yes/no queries concerning structural features, and therefore fingerprints could represent high-dimensional chemistry spaces (10^2–10^6 bits per string). While these were useful in identifying similar compounds within large databases, such algorithms did not adequately address issues of dissimilarity[320]. Moreover, measurement of the relative distance between compounds in chemistry space is not applicable to other diversity assessment tasks which require absolute coordinates for each compound.

5.2.1 BCUT chemical indices

A more sophisticated, 'cell-based' approach, which operates in terms of absolute coordinates, involves the division of each metric axis into discrete 'bins' that generate cubic cells in multi-dimensional space. Since the number of cells increases at a rate of 2^n, where n is equal to the number of dimensions it is only practical to select several (<10) metrics[320]. Empirically, it was found that whole-molecule descriptors, such as molecular weight, solubility, and pK_a, are not useful in this situation as they do not contain sufficient information concerning substructural diversity. Instead metrics related to biomolecular interactions, such as atomic charges and polarisabilities, and hydrogen bonding potentials, coupled to 3D structural information proved to be much more effective, and has given rise to the current 'BCUT'† approach[320].

The software package *DiverseSolutions*‡ employs the BCUT approach to explore compound diversity. Chemistry space coordinates are calculated for a library of compounds according to several metrics. It is important to then examine the distribution of compounds along each axis as some metrics will be more effective than

*For a recent review of this field see *Nature* magazine, 'Insight' supplement, December 16th 2004

†The acronym BCUT is derived from the author that inspired this approach, F.R. Burden, and the two institutions that used his ideas to develop the method further, namely the Chemical Abstracts Service (CAS) and the University of Texas.

‡www.tripos.com/sciTech/inSilicoDisc/comboChem/dvs.html

others in distinguishing between compounds. It would be of little advantage to use a metric that clustered most of the compounds within a small region of space. DiverseSolutions includes an algorithm to select six combinations of metrics that generate compound distributions as close as possible to an idealised model of diversity (uniform dispersal in chemical space). The choice of metrics is validated further by analysing the distribution of known ligands for the given receptor, which are expected to be clustered in chemistry-space.

Once a meaningful classification of compounds has been achieved, *DiverseSolutions* enables the user to explore diversity within the library, select representative compounds from each cell, and to compare different libraries. It is expected that a given target will interact with ligands in a discriminatory manner, with preference for certain chemistry and shape. *DiverseSolutions* includes a nearest-neighbour algorithm that reports the n closest compounds to each known ligand within the given chemical space dimensionality. The program was employed in this manner to select fragments that are chemically similar to known SH2 domain ligands, as described in Section 5.4 below.

5.3 Identification of pY mimetics using FlexX

FLEXX was used to analyse the interaction between the *v*-Src SH2 domain and a large library of fragment compounds. Initially, a series of known ligands was used to test the software and optimise virtual screening conditions. An Octane Silicon Graphics Inc. (SGI) machine was used to perform docking calculations within the context of the molecular graphics package SYBYL[§].

5.3.1 Selection and preparation of protein structure

Fragment docking with *v*-Src SH2 required selection of a suitable target structure. As stated previously, the *v*-Src SH2 domain structure has been solved in the peptide-bound (1SPS) and 'peptide-free' forms (1SPR), and now also in the apo solution state. Whilst it would be preferable to use unliganded protein structures in docking calculations, the absence of ligand-induced structural changes in the protein may adversely affect the predicted interaction of such ligands, resulting in mediocre scores. The alternative is to use a bound-state structure, as this at least reflects the protein structure in the presence of a known ligand. However, bound-state structures will be subtly different for every ligand, yet these remain inaccessible in FLEXX calculations. On balance, it is argued that docking into the complex structure is more likely to be successful in identifying novel ligands than using the uncomplexed state.

[§]www.tripos.com/sciTech/inSilicoDisc/moleculeModeling/sybase.html

Although the peptide-bound and peptide-free *v*-Src SH2 crystal structures are structurally very similar (C_α rms deviation = 0.56 Å), there are differences in side chain positions, including important binding site residues (Figure 5.2). The most significant discrepancy is for R12, which forms two charge-assisted hydrogen bonds with phosphate oxygen. In the apo state, the R12 guanidinium moiety (C_ζ atom) is rotated away by 2.0 Å. Smaller differences exist for the side chains of T36, C42 and K60. Deviation from optimal bound-state geometries could potentially generate unfavourable docking scores for known binders, and instead highlight a different set of fragments that may not in fact display real affinity. In view of these observations, the peptide-bound form of *v*-Src SH2 was selected as the docking target.

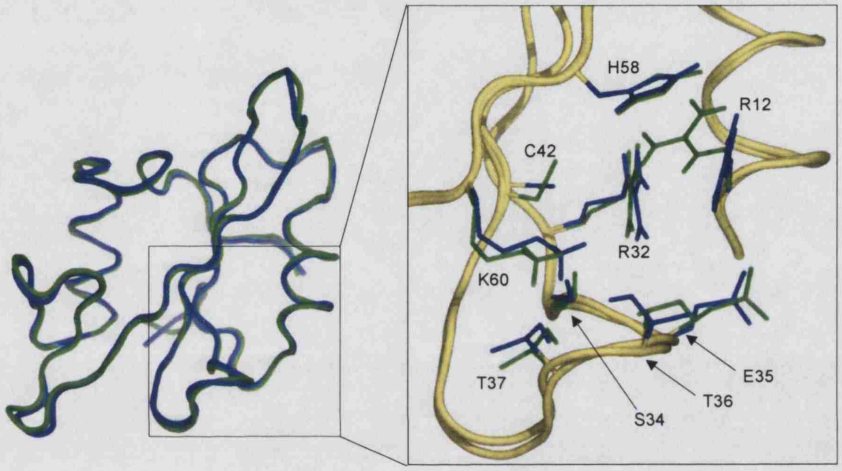


Figure 5.2 *v*-Src SH2 structures fitted using C_α atoms; 1SPR peptide-free (blue) and 1SPS peptide-bound (green). Important side chains in the pY binding pocket are shown as sticks.

The *v*-Src SH2 peptide-bound structure file contains three SH2 domains per asymmetric unit (Mol A–C), two of which (Mol A & C) form mutual intermolecular contacts within the crystal. Coordinates for the isolated molecule (Mol B) were therefore selected as the docking target and extracted to a single file. Atoms were protonated as necessary using SYBYL. To ensure this did not lead to unrealistic atomic clashes within the structure, the energy of the protein was calculated using the TRIPOS force field, yielding a low value within the expected range for a 'clean' structure. The active site for docking calculations was typically defined as any protein atom within 6.5 Å of the phosphotyrosyl side chain from the peptide. This included 171 heavy atoms from 20 residues[¶].

[¶]I10–R13, S15, R32–T37, A40–L43, K57–I61

5.3.2 Assessment of FLEXX docking capabilities

In order to probe the response of v-Src SH2 to fragment docking calculations and optimise experimental conditions, a series of phosphate-based fragment compounds were prepared with SYBYL and used in several important validating tests. These focussed on how well FLEXX could reproduce experimental data concerning binding specificity, affinity, and charge states.

5.3.2.1 Specificity

Docking calculations are typically restricted to a small area of interest on the protein surface. This serves to filter out compounds with predicted affinity for distal sites as well as reduce the overall mathematical complexity. However, compounds shown to interact with the main site may still be non-specific binders that also interact elsewhere on the protein surface *in silico* and/or *in vitro*. With the exception of SAP[‖], SH2 domains are highly specific towards phosphotyrosine-based peptides[110]. Other studies have indicated that SH2 domains, including that of Src, interact with peptides bearing multiple phosphotyrosyl epitopes[278, 131]. However, the NMR titration experiment described above (Section 4.1.2) demonstrated that phosphotyrosine amino acid only interacts with the positively-charged binding pocket of v-Src SH2. In support of this, thermodynamic analysis of pY binding reported a stoichiometry of 1:1[106].

The ability of FLEXX to replicate these *in vitro* studies was tested by docking phosphotyrosine against the entire v-Src SH2 domain surface. The amino and carboxyl groups of phosphotyrosine were methylated in order to remove undesirable zwitterionic charges, and simulate its environment within a larger peptide. The docking results indicated a high degree of accuracy and precision in ligand placement (Figure 5.3). Moreover, the position of phosphate and ring atoms were very close to those observed in the structure of the peptide complex. The pattern of chemical shift changes induced by pY (Section 4.1.2) suggested that this ligand also adopts a peptide-like orientation *in vitro*. The demonstration that FLEXX can model the correct interaction between v-Src SH2 and pY permits restriction of the docking target to the main binding site and improves the credibility of future docking results for this domain.

[‖]SAP SH2 makes a larger number of interactions with target peptides which enable it to bind non-phosphorylated tyrosine sequences[321, 322].

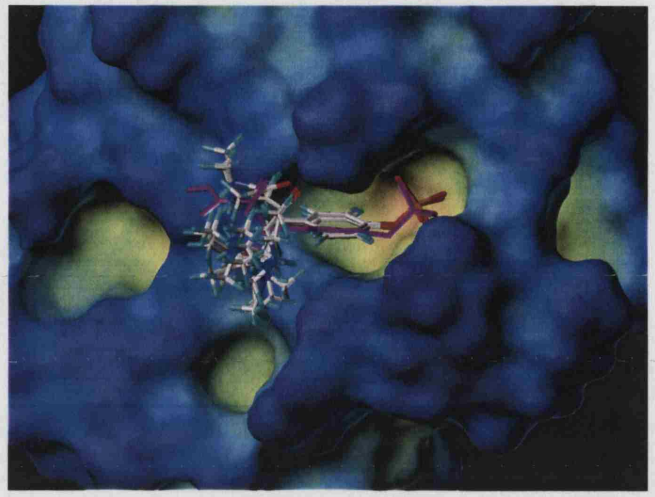


Figure 5.3 Testing the specificity of pY binding *in silico*. This figure indicates the result of 30 FLEXX docking attempts (shown as multi-coloured sticks) in comparison to the position of pY within the known crystal structure (purple sticks). The *v*-Src SH2 domain surface is coloured according to cavity depth.

5.3.2.2 Fragment size

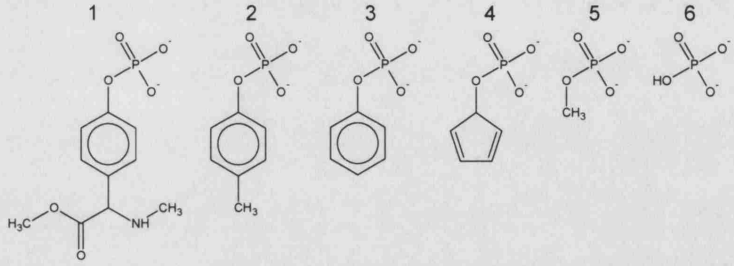
Consistent placement of small fragments is sometimes not achievable in docking exercises, in part due to the large number of potential binding sites and modes. To test the influence of fragment size with the *v*-Src SH2 domain, a group of phosphate-based fragments (1-6, Figure 5.4) was designed and subjected to docking against the phosphotyrosine pocket. Each fragment was docked thirty times; FLEXX results in Figure 5.4 represent the highest scoring attempt. All six ligands were assigned highly favourable FLEXX scores, although this clearly arose from the common element of each fragment – phosphate. This concurs with the general conclusion from SH2-related drug design research that phosphate is the source of the majority of the binding energy. Attempts to dock small, non-polar fragments (e.g. methane, ethane, ethene) either failed completely, or generated very low FLEXX scores (data not shown).

Upon examination of the highest-scoring docking positions it was found that certain fragments interact in a slightly different manner compared to pY (Figure 5.4). Across the six ligands, the positions of two phosphate oxygen atoms were conserved, however in 3, 5, and 6, the phosphate group was rotated such that the other two oxygens did not interact with the domain. This caused side chains of 3 and 5 to be orientated differently. It is not known whether this behaviour is observed *in vitro*. Despite these differences, FLEXX appears to be able to perform docking calculations with very small fragments, such as phosphate, as long as the compound contains sufficient polar atoms capable of hydrogen bonding with *v*-Src SH2. The preference for such interactions has been demonstrated regularly throughout the history of

SH2-related drug design.

5.3.2.3 Fragment charge

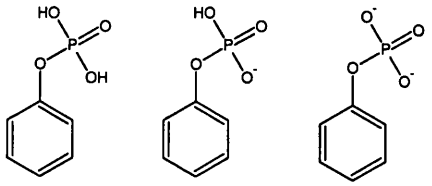
On binding to *v*-Src SH2, the pK_a of phosphotyrosine is reduced, leading to full deprotonation ($-OPO_3^{2-}$) at pH 6.0 (see Section 4.3.7). Since FLEXX does not attempt to control the protonation state of ionisable groups according to the microenvironment of the ligand, it may be important to ensure that such groups are appropriately defined prior to analysis. In order to investigate whether FLEXX score is strongly dependent on the charge state of fragments, three phenylphosphate compounds (7-9)



Fragment	FLEXX docking score
1 (phosphotyrosine N/C-methylated)	-42.8
2 (4-methyl-phenylphosphate)	-41.1
3 (phenylphosphate)	-40.7
4 (cyclopenta-2,4-dien-1-yl-phosphate)	-40.3
5 (methylphosphate)	-39.4
6 (phosphate)	-39.0

Figure 5.4 Investigation of the effect of fragment size on FLEXX docking score. The highest-scoring docking positions (multi-coloured sticks) are shown for each fragment at the bottom of the figure. The orientation of pY within the peptide-bound crystal structure is shown as purple sticks. The *v*-Src SH2 surface is coloured according to cavity depth.

were prepared as neutral, single-, and double-negatively charged species, and docked into the pY pocket of v-Src SH2 (Figure 5.5). FLEXX generated significantly more favourable interaction scores for the anionic fragments. This is to be expected since the pY pocket has a relatively high local positive charge and because the additional protons of 7 reduces the number of lone pairs available for hydrogen bonding and provides steric hindrance to binding. Interestingly, there is a slight preference for the singly charged state, which does not correlate with *in vitro* results obtained during this project, or with previously determined pK_a values[256, 105]. In conclusion, the charge of a potential ligand can have a significant effect on the final docking outcome (ΔG_{io} in Equation 5.1), which highlights the need to specify ionisation states correctly in future calculations.



7 **8** **9**

Fragment	Charge	FLEXX score
7	0	-17.1
8	-1	-39.3
9	-2	-38.3

Figure 5.5 Investigation of the effect of phenylphosphate fragment charge on FLEXX docking score.

5.3.3 Docking of known binders

The performance of FLEXX was further investigated by docking a series of phosphotyrosine mimetic fragments (n = 44), identified in the literature as displaying affinity (or IC_{50} value) for an SH2 domain similar to that of phosphotyrosine. In the past five years, there have been several publications in which phosphotyrosine replacements have been assayed in the fragment form (e.g. see refs[61, 180]). Prior to this, these fragments were generally tested in the context of a larger, high affinity scaffold, such as a peptide or peptidomimetic (e.g. Ye *et al.*[323]). In these latter cases, the phosphotyrosyl replacement group was extracted and treated as an independent binding fragment. Efforts were made to include phosphotyrosine replacements devoid of phosphate, resulting in a range of alternative functionalities, as shown in Table 5.1.

The 44 selected fragments, hereafter termed 'known binders' (see Figure 5.6), were constructed in SYBYL, energy-minimised, and ionised using *Pipeline Pilot**.

*www.scitegic.com/products_services/pipeline_pilot.htm

Fragment functionality	Number in library
Carboxylate	14
Nitro	1
Phosphate	16
Phosphonate	9
Sulphate	1
Sulphonate	2
Thiophosphate	1

Table 5.1 Distribution of side chain functionality within the 44-member library of SH2 ligands.

Each fragment was subjected to thirty docking attempts into the phosphotyrosine pocket of *v*-Src SH2. The resultant FLEXX scores (Figure 5.7) indicated that the most successful mimetics contained either phosphate or phosphonate groups. Fragments bearing carboxylate moieties achieved slightly lower scores and sulphur-containing compounds performed relatively poorly. Whilst this pattern of results agrees with the overall body of SH2 drug design literature, a quantitative comparison between FLEXX score and empirical *in vitro* affinity/IC_{50} was not possible as the latter values were derived from a range of SH2 domains and analytical techniques.

The difference between SH2_001 (phenylphosphate) and SH2_020 (phenylthiophosphate) is conservative replacement of oxygen by sulphur (Figure 5.8). However, this caused an unfavourable positive increase in FLEXX docking score (-41 to -14), which can be explained by considering the fact that the atomic volume of sulphur is 66 % greater than that of oxygen[†]. As Figure 5.8 indicates, the docking position of SH2_020 is significantly different to SH2_001, which adopts a phosphotyrosine-like position. The sulphur atom is solvent-exposed and positioned at the open end of the pY pocket.

It was therefore established that FLEXX could identify known binders, and broadly distinguish between higher affinity (phosphate) and lower affinity (carboxylate) fragments. This provides a high level of confidence in future docking calculations, including large scale virtual screening.

5.3.4 Docking of fragment library

Having validated and optimised the application of FLEXX to docking against the SH2 domain, a large library of fragment compounds was screened. Compound structures were obtained from the Available Chemical Database (ACD), which at the time of investigation (2003) consisted of over 170,000 compounds. A total of 12969

[†]Based on atomic radii

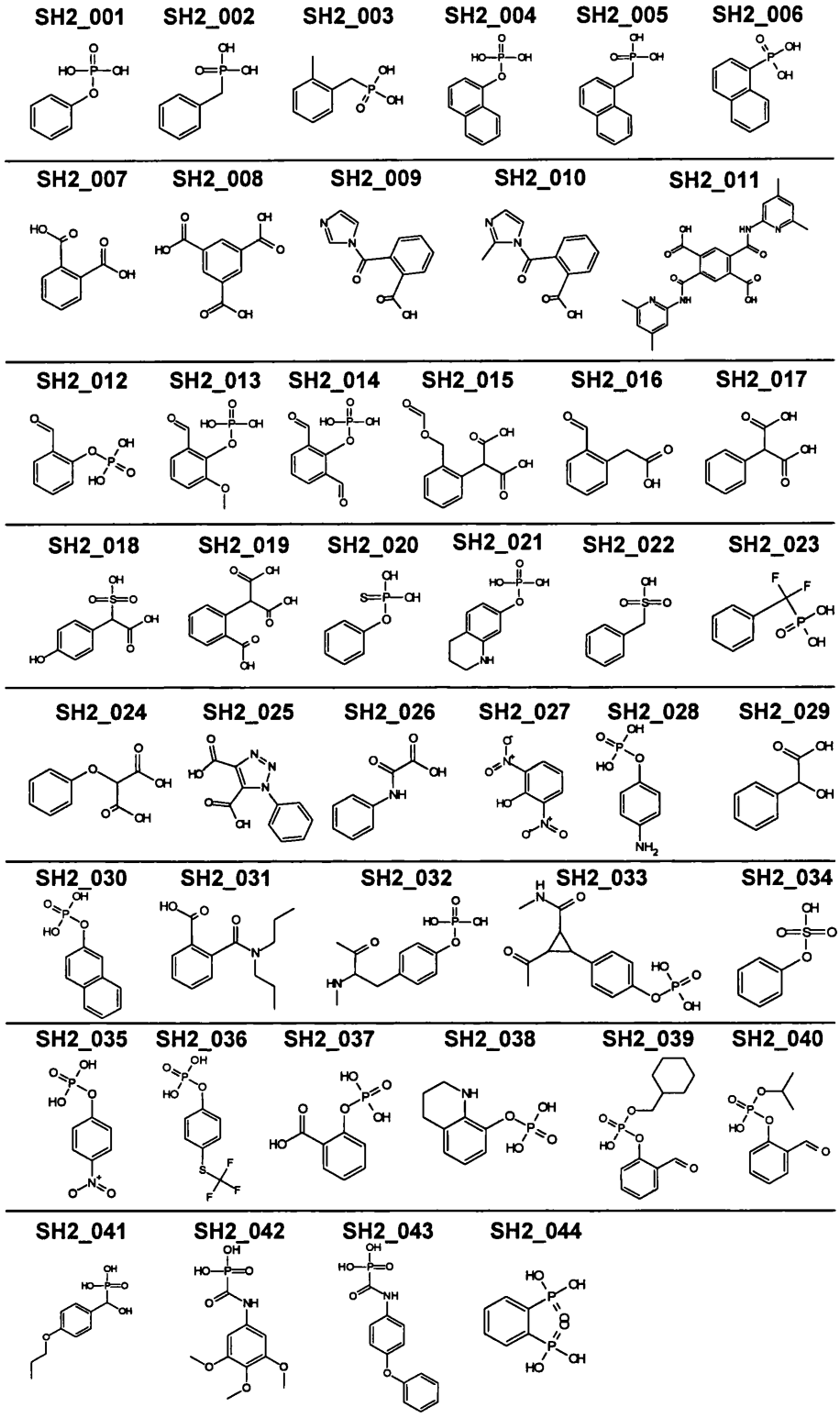


Figure 5.6 Structures of the 44 known binders used to assess fragment docking with *v*-Src SH2

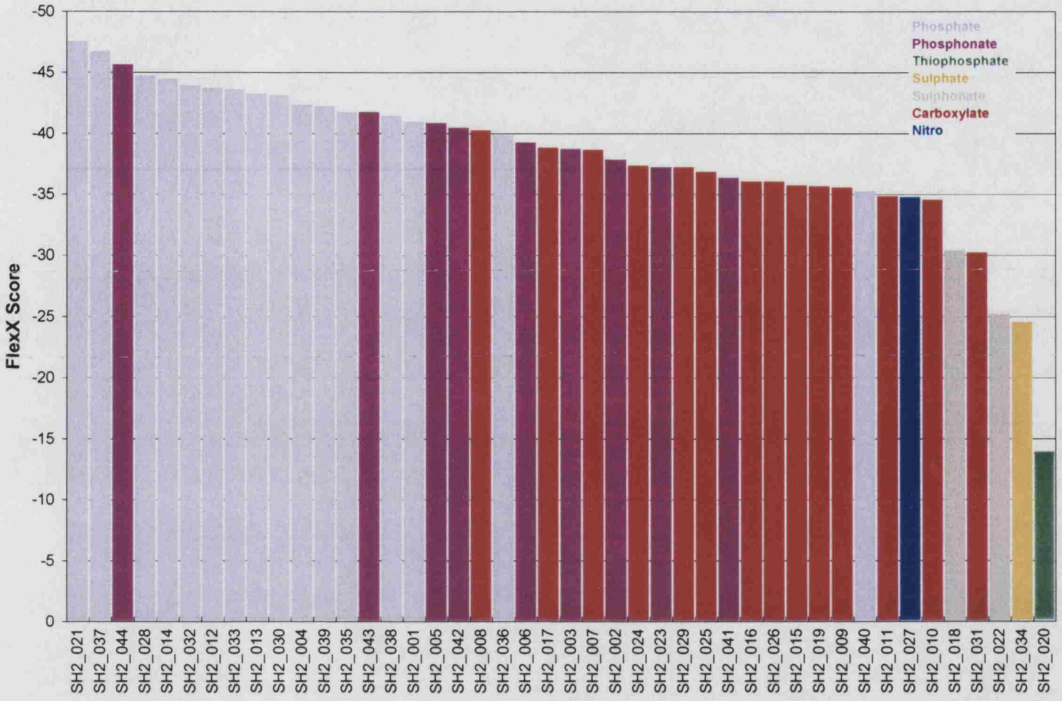


Figure 5.7 Ranked FLEXX docking scores for the 44 known binders, coloured according to functional group type.

fragments were extracted from this database by applying a series of filters[‡] (Figure 5.9). Selection criteria focussed on excluding compounds that were inappropriate for *in vitro* screening and including structures that could be considered as fragments. Such criteria were intermediate between Lipinski's 'rule of five'[49] and the fragment-based 'rule of three' proposed by Congreve *et al*[51§]. Selected compounds had an average molecular weight of 189 Da, and consisted of 23.6 atoms, 24.0 bonds (3.7 of which were rotatable), and 1.5 ring systems.

Fragment structures were stored in the SMILES[¶] code format, which describes in a single string the atom type, charge and bonding order of a molecule, with an associated reference key. Docking calculations required conversion of this SMILES codes into 3-dimensional energy-minimised structures. This was performed using CONCORD[‖], a program within SYBYL. The pK_a of fragment atoms was predicted using a module from the *Pipeline Pilot* chemoinformatics suite and ionisable groups were protonated appropriately for pH 6. Importantly, this resulted in full ionisation of phosphate groups (OPO_3^{2-}). The secondary pK_a of phosphotyrosine in free solution is approximately 6.2 however as described in the previous chapter, binding

[‡]This was done by Dr Will Pitt, Celltech

[§]It should be noted that at the time of compound selection, the 'rule of three' had not yet been published.

[¶]SMILES = simplified molecular input line entry system

[‖]See Pearlman[324] or www.tripos.com/sciTech/inSilicoDisc/chemInfo/concord.html

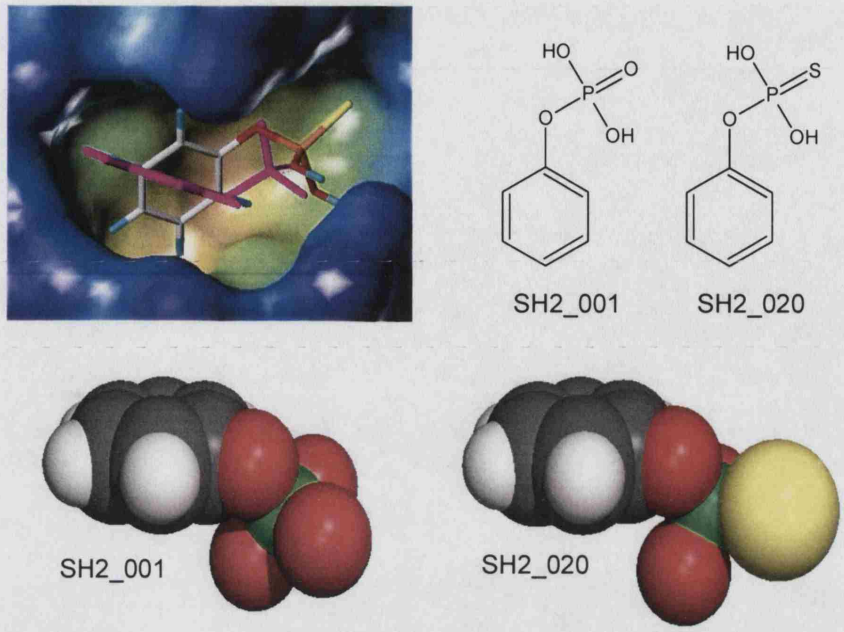


Figure 5.8 Comparing the docking behaviour of two similar fragments. *Top row* | Chemical structures and docking positions of SH2_001 (purple sticks) and SH2_020 (multi-coloured sticks). *Bottom row* | Atomic sphere representation of SH2_001 and SH2_020 indicates the difference in size between the oxygen and sulphur atoms, coloured red and yellow, respectively.

of phosphotyrosyl peptides to SH2 domains involves a lowering of this value by 1-2 units, which favours the double-negative charge state. In this case it was fortunate that pK_a estimation could be validated by experimental data.

The 12969 ACD compounds that fulfilled the selection criteria were subjected to docking calculations against the pY pocket of the *v*-Src SH2 domain. For comparison, the 44 known binders were also included within the list of fragments. This large computational task required 80 hours of CPU time on a standard Octane SGI workstation. Over 99 % of the ACD compounds were docked successfully (125 completely failed), generating FLEXX scores between -48.9 and +2.7 and mean ± standard deviation values of -16.6 ± 9.5. The histogram of FLEXX scores (Figure 5.10) indicated a complex distribution, with two peaks centered -30 and -10.

The docking software packages currently in use differ in terms of the methodology used to score the free energy of an interaction. Celltech uses FLEXX for virtual screening, but combines a number of different scoring methods in the analysis of each compound. This is implemented by the Consensus Score (CSCORE) program, integrated to SYBYL, that assesses ligand interactions using five alternative models. Three of these are based on scoring routines from GOLD, DOCK and FLEXX; the other two utilise statistical approaches:

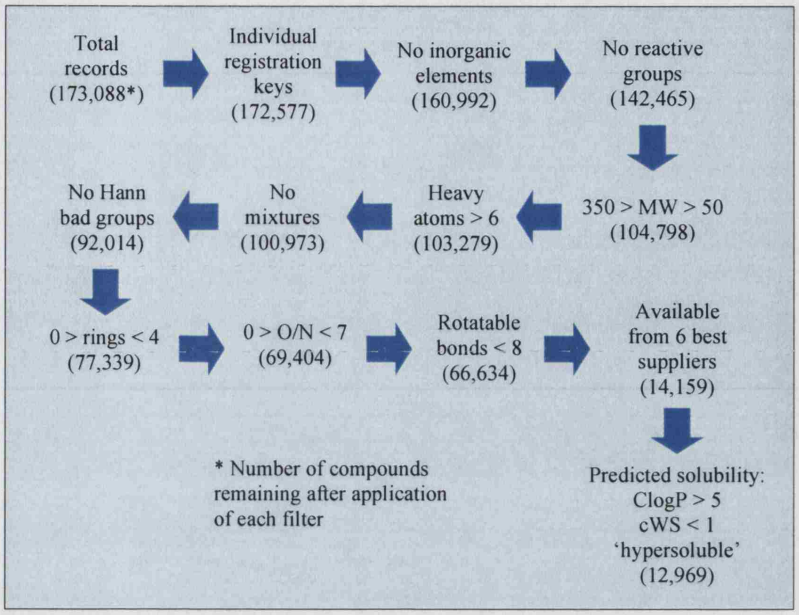


Figure 5.9 Sequential filtering of the ACD to isolate fragments useful for virtual screening

- **G-SCORE** – A GOLD-like function that is good at quantifying hydrogen bond contributions but has difficulty predicting hydrophobic interactions[309].
- **D-SCORE** – This DOCK-like function includes electrostatic and hydrophobic terms but excludes entropic information[325].
- **F-SCORE** – The FLEXX score features equations that divide H-bond, salt-bridge and non-polar interaction distances into discrete bins. An energy is assigned to each interaction based on which bin it occupies. Entropic and enthalpic terms are also considered, as well as the effect of buried surface area and the number of rotatable bonds[311].
- **CHEMSCORE** – This method was derived from a training set of 82 receptor-ligand complexes. It assesses contributions to binding energy from lipophilic, hydrogen bonding, and metal-ligand interactions and loss of ligand flexibility[326].
- **PMF-SCORE** – This function is based upon ligand-receptor atom-pair interaction potentials. Such potentials were derived from crystallographic data in the PDB[231] and hence the results are statistical in nature rather than empirical[327].

CSCORE was used to analyse the highest-scoring interaction between each ACD fragment and the pY pocket of *v*-Src SH2 (Figure 5.11). The results within each scoring function were normalised across all compounds in the following exemplary manner:

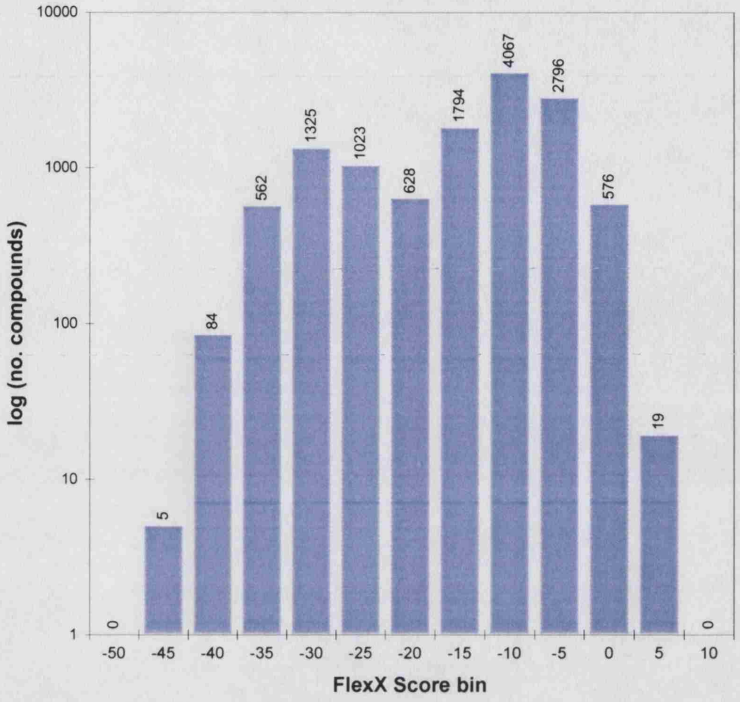


Figure 5.10 Histogram showing the distribution of FLEXX scores obtained for the 13013 fragment compounds. Mean ± standard deviation of data was -16.6 ± 9.5.

$$\text{Normalised G-SCORE value } (G_i^n) = \frac{(G_i - \overline{G})}{\sigma_G} \tag{5.2}$$

where G_i, $\overline{G}$ and σ_G correspond to the G-SCORE value of the i^{th} fragment, the mean of all docked fragments, and the standard deviation of G-SCORE values, respectively. The five normalised scores per compound were subsequently averaged to give a combined score. Compounds were ranked according to this final score, the distribution of which is shown in Figure 5.12. The purpose of this statistical data processing step was to allow a rational quantitative comparison of combined fragment docking scores. Since each scoring method generates a different range of data, the scores must be internally normalised prior to comparison with data from other methods.

5.3.5 Assessment of scoring algorithms

The performance of FLEXX and CSCORE can be assessed by examining the ranking positions of the 44 known binders using an 'enrichment plot'. It is expected that these ligands would achieve high ranking positions in the list of scores. Confidence in the results would be diminished if it was found that a large proportion of known binders achieved relatively low scores. To analyse the location of known binders

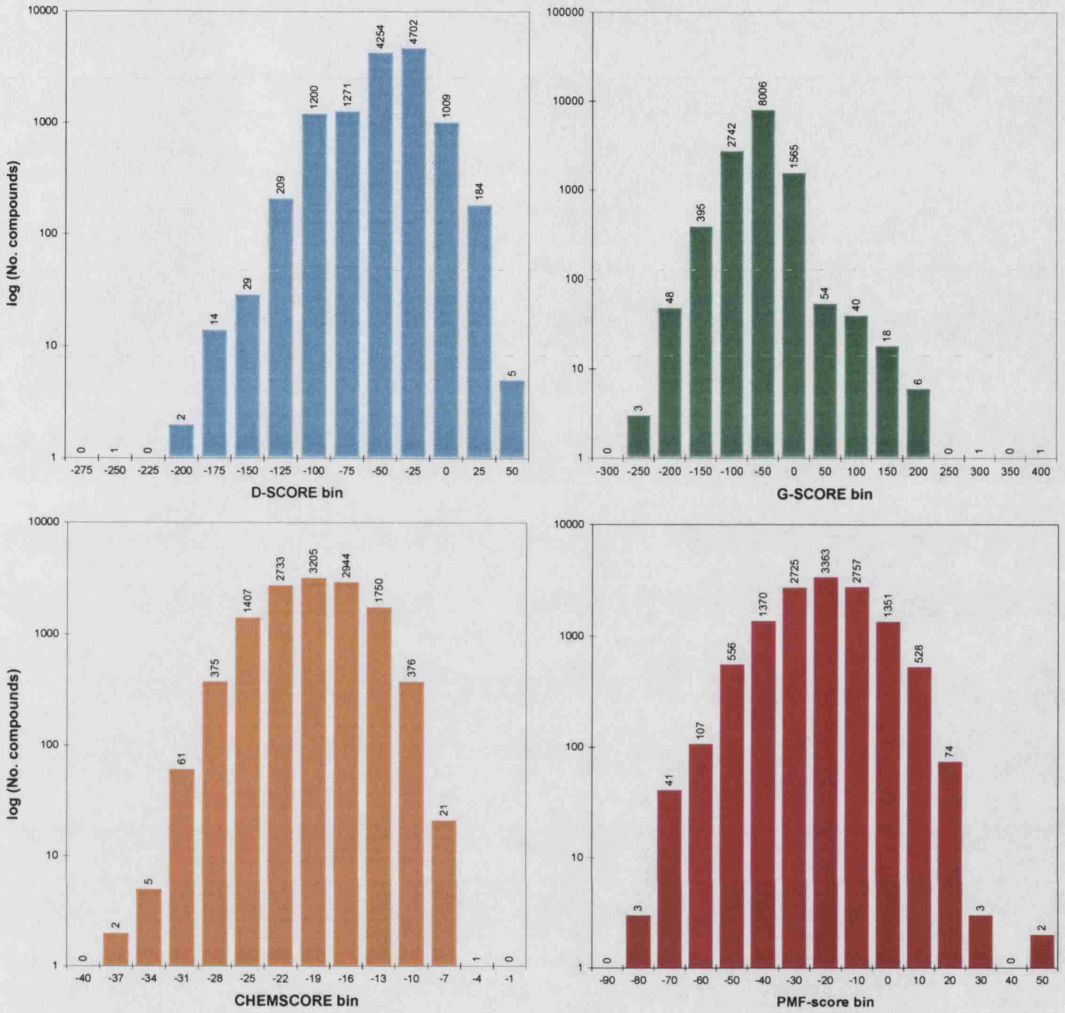


Figure 5.11 Histograms showing the distribution of values obtained for each CSCORE function, except the F-SCORE, which is identical to that of FLEXX (see Figure 5.10). In each case, the more negative the score, the more favourable the interaction. Mean ± standard deviation values for each data set were as follows: D-SCORE = -57.0 ± 29.2; G-SCORE = -84.6 ± 34.6; CHEMSCORE = -20.3 ± 4.2; PMF-SCORE = -25.3 ± 15.0.

within the larger data set, a threshold score is chosen and the number of known binders achieving a more favourable (more negative) score counted. Threshold values are then progressively increased until the entire data set is selected, which, by definition contains 100 % of the known binders. If scoring was random, on average, the known binders would be evenly distributed throughout the list. Alternatively, an 'ideal' scoring system might give top ranking positions to known binders. The location of the 44 known binders was analysed in this way, comparing the observed data to both 'random' and 'ideal' distributions (Figure 5.13).

It was found that 100 % of the known binders occur within the top 4000 compounds, however 26 (59 %) were ranked among the top 100. The observed distri-

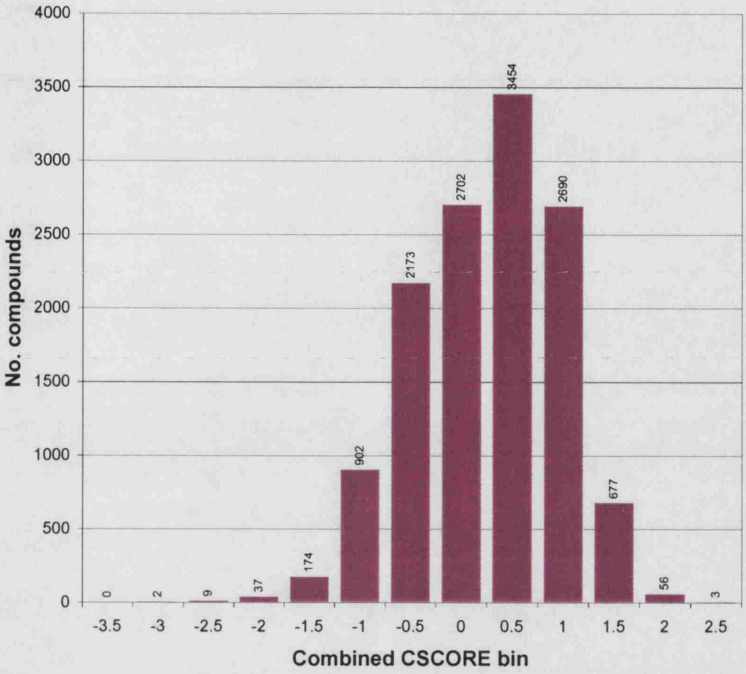


Figure 5.12 Histogram showing distribution of combined scores of the 12888 ACD compounds and 44 known binders that were successfully docked into the pY pocket of *v*-Src SH2.

bution resembles more closely the ideal data where known ligands are given top ranking scores and thus it is reasonable to assume that fragments with similarly favourable docking scores might display affinity for *v*-Src SH2 *in vitro*.

5.4 Identification of pY mimetics using *DiverseSolutions*

Chemical similarity indices were obtained for the 44 known binders and 12969-member ACD fragment library using *DiverseSolutions*. This required prior conversion of SMILES strings into energy-minimised, and correctly ionised 3D structures, performed using CONCORD, SYBYL, and *Pipeline Pilot* as described above. The four chemistry space metrics that generated the highest degree of fragment dispersal were selected automatically. These were 'hydrogen acceptor,' 'hydrogen donor', 'Gasteiger charge'[328], and 'polarisability'**. The set of known binders were subsequently used to perceive the most relevant metric, which was 'hydrogen acceptor'. The central pharmacophore of SH2 domain ligands is phosphate, an anionic, multiple hydrogen bond acceptor. *DiverseSolutions* was able to detect this pattern within the set of known binders.

**Polarisability is the ability of an ion to distort the electron cloud of another. Covalent bonding occurs where there is a large amount of polarisation, resulting in sharing of electrons between elements.

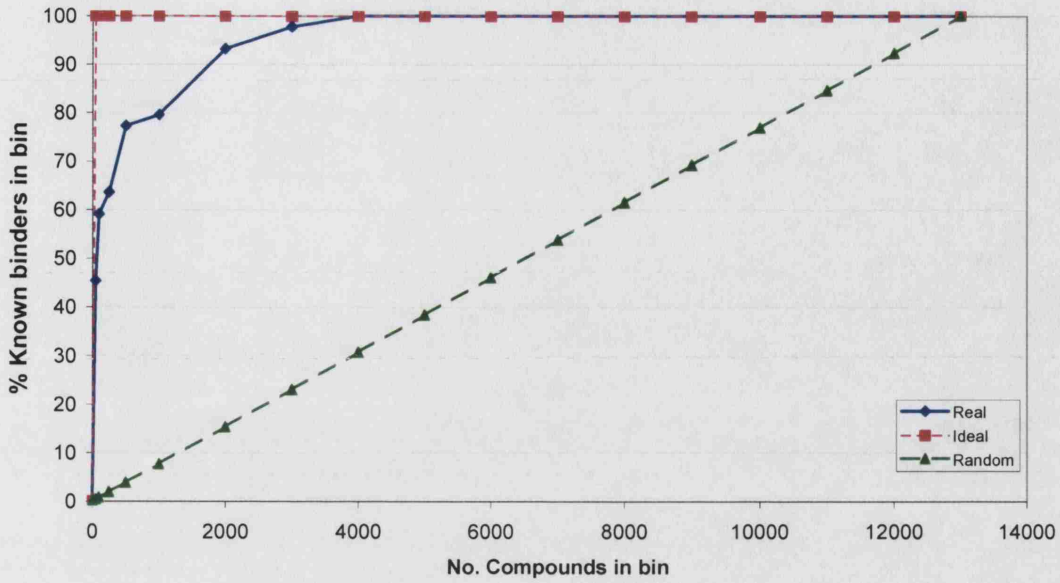


Figure 5.13 The enrichment plot: An assessment of the ranking positions of the 44 known binders within the list of combined CSCORE values. The threshold ranking position is gradually lowered, an increasing number of known binders are encountered. This is compared with ideal and randomised data sets.

The distribution of compounds in chemical space (Figure 5.14) was analysed using *Spotfire DecisionSite*[††] and *Microsoft Excel*. All coordinate values were normalised to the range 0 to 10 in order to facilitate this comparison. The set of known binders was not found to be tightly clustered with a single region of chemical space, however some grouping was observed for fragments containing common functionalities, for example, phosphates or carboxylates. The ~13,000 compounds were not evenly distributed throughout chemical space indicating a degree of redundancy within the data. Certain regions were devoid of fragment coordinates suggesting that the diversity in the library of compounds was sub-optimal.

DiverseSolutions was used to identify the 10 nearest neighbours to each known binder, according to the four metrics that were shown to provide the most uniform distribution of fragments. Analysis of this subset of compounds indicated a range of fragments that were surprisingly diverse in terms of structure, yet arguably similar to known binders. In many cases, it was not immediately obvious how the identified compounds would interact with the *v*-Src SH2 domain binding site. It was decided that a combination of results from fragment docking and nearest neighbour analysis would be the most efficient method of identifying novel fragments with likely binding affinity for the SH2 domain.

[††]www.spotfire.com/products/decision.asp

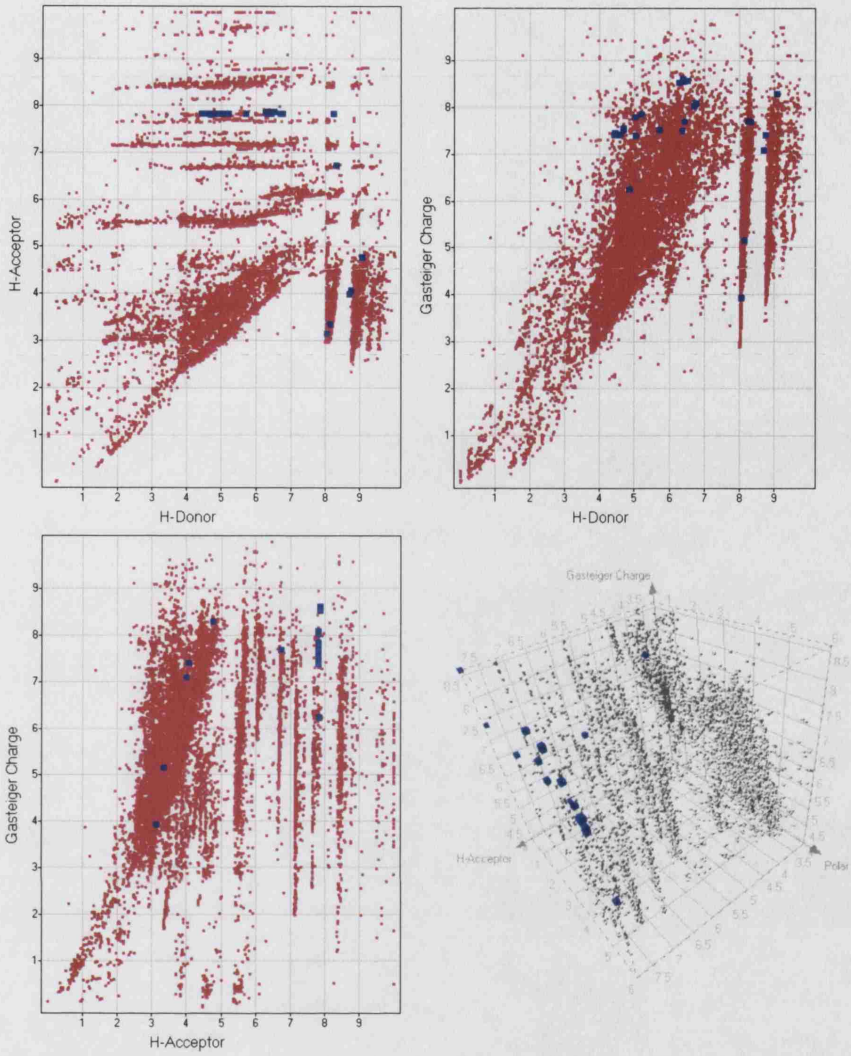


Figure 5.14 Sample plots showing the distribution of known binders (blue) and ACD fragments (red/grey) in chemical space in two- and three-dimensions according to various metrics.

5.5 Selection of fragment compounds for *in vitro* screening

A total of 76 unique fragments that were chemically similar to the 44 known binders, and had achieved high ranking docking scores were identified (Figure 5.15) and considered in greater detail. The predicted binding orientation of each fragment was analysed according to the following set of criteria:

1. Does the fragment make several of the most conserved contacts with the core pY pocket residues (R12, R32, S34, E35, T36, C42, K60)?
2. Does the fragment make any additional, but plausible, contacts that might confer specificity to the *v*-Src SH2 domain?

3. If present, do the hydrophobic parts of the fragment interact with the protein surface or are they fully solvent-exposed?
4. Does the fragment fill the pY pocket in terms of depth and/or breadth?
5. Are there any obvious chemical modifications that could be made to improve compound affinity or specificity?
6. Could slight alteration of the SH2 domain structure accommodate lead to more optimal contacts (i.e. induced fit)?

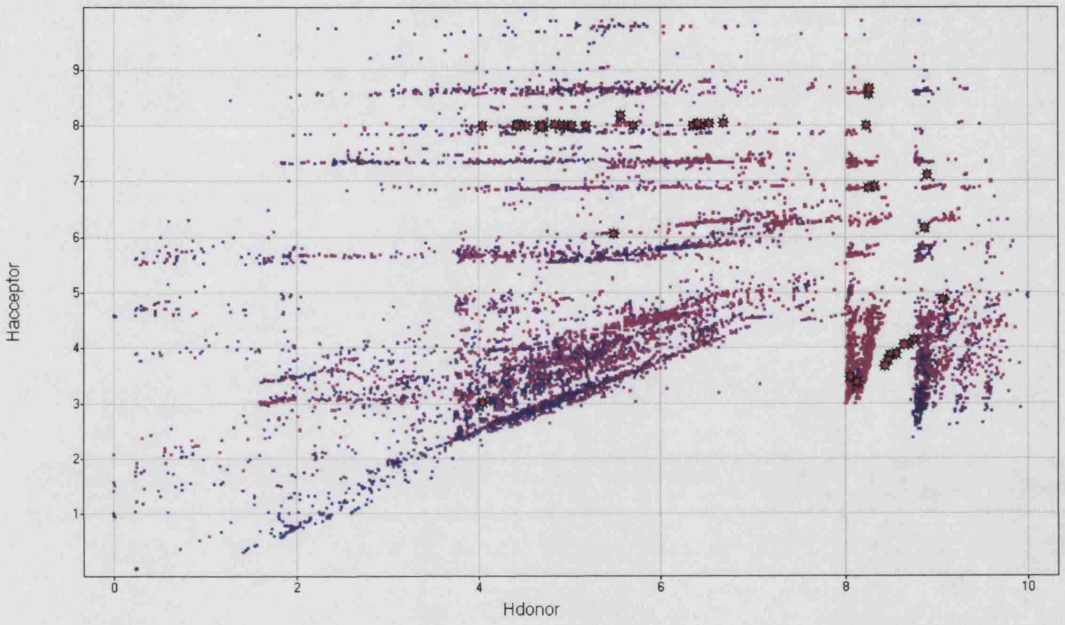


Figure 5.15 The distribution of ACD fragments (dots) and known binders (red stars) in chemical space according to metrics related to hydrogen bonding. The data points are coloured using a blue→red gradient describing positive→negative (poor→good) combined docking scores associated with each fragment.

Using these questions to filter out unlikely ligands, a final list of 14 compounds was prepared (Figure 5.16 & Table 5.2). Since the NMR screening experiment described in the next chapter required a reporter ligand containing fluorine, it was necessary to include at least one such compound in this final list. Unfortunately, due to various restrictions[‡‡] four compounds were inaccessible at this stage (MFCD numbers 1863005, 63405, 671777 & 135087). Two known binders – phenyl phosphate and 1-naphthyl phosphate – were also purchased for use in screening studies. During the course of this investigation, fluorophosphate (FPO_3) also emerged as a potential weak-binding ^{19}F-containing molecule that could be used in the NMR screening experiments. This compound was not accepted as a valid fragment during the initial selection of compounds (Figure 5.9).

[‡‡]For example, government restrictions, non-availability, high cost

MFCD No.	Docking score	MW (Da)	clogP	PSA (Å^2)	Rotatable bonds	H-bond acc.	H-bond don.	Ring count	Chiral centers	Atom count
Novel compounds										
00036757	-2.39	325.5	1.79	172.4	3	4	4	2	0	21
00672514	-1.86	298.8	2.85	95.1	7	4	1	1	1	37
00063405	-1.14	331.7	-0.09	169.0	4	3	4	3	1	35
00004447	-1.80	227.2	1.80	107.7	4	3	1	2	0	28
00671777	-2.14	333.2	3.12	109.4	7	3	3	2	1	39
00079679	-2.15	320.4	3.17	124.7	7	4	1	2	1	40
00057611	-1.86	278.2	-1.36	178.1	4	5	1	2	0	28
00049223	-1.92	296.1	3.23	99.8	4	3	1	2	0	28
00135087	-1.72	307.4	2.25	214.7	5	5	1	3	1	31
00005625	-1.86	188.2	0.56	174.5	2	3	1	2	0	20
00800300	-2.29	309.1	0.29	210.9	4	6	3	2	0	25
01863377	-2.80	330.2	2.08	212.0	6	6	3	2	0	30
01863005	-2.21	290.1	0.19	199.0	3	6	3	2	0	24
Mean	-2.04	297.3	1.55	160.0	5	4	2	2	-	29
Known binders										
Phenyl phosphate	-1.81	173.1	0.14	146.1	3	4	3	1	0	18
1-naphthyl phosphate	-2.20	223.1	1.31	139.1	3	4	3	2	0	24

Table 5.2 Properties of novel compounds identified by docking and chemical similarity indexing, as well as two known binders. Docking score, combined docking score; cLogP, calculated log[solubility in water/solubility in octane]; PSA, polar surface area.

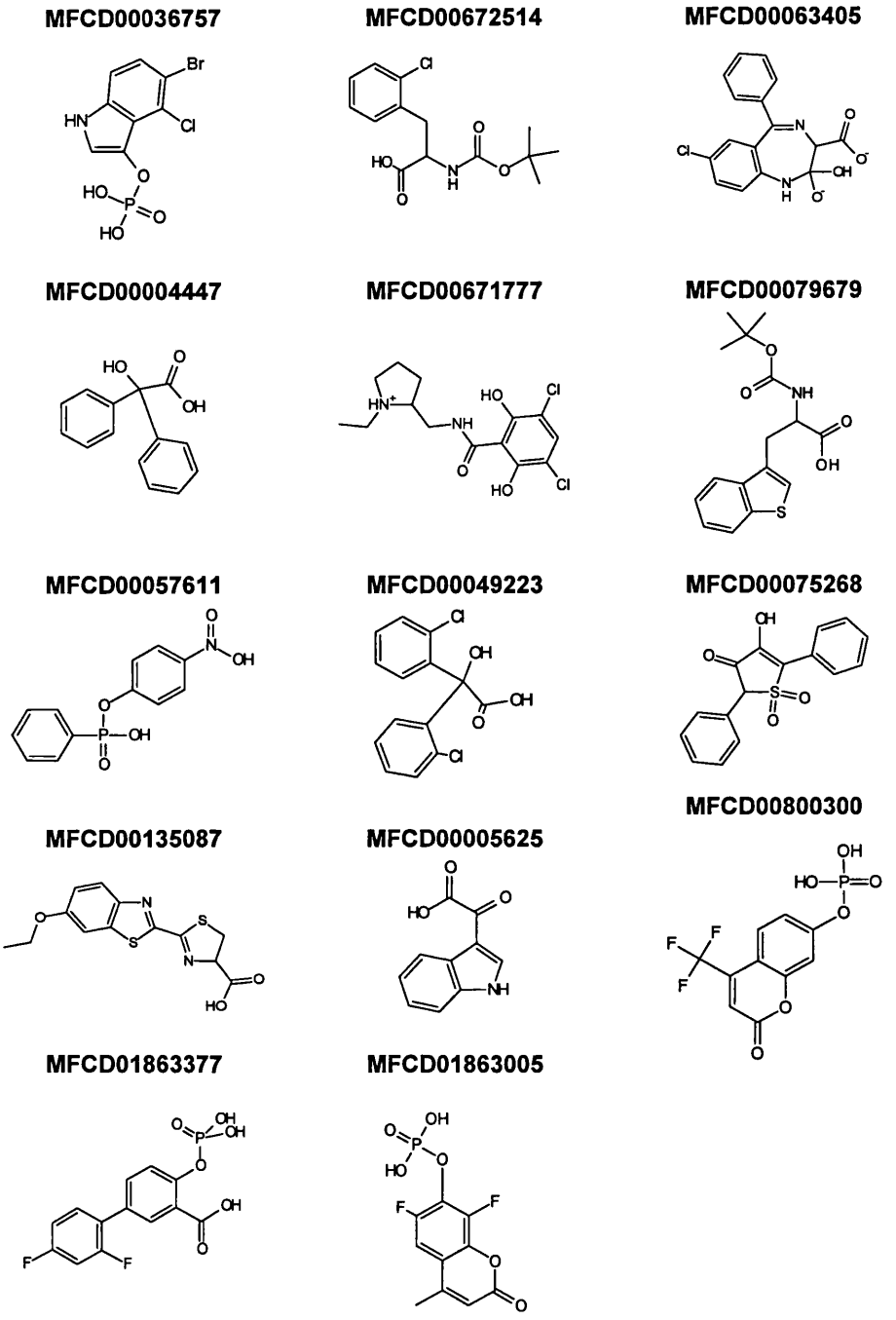


Figure 5.16 Chemical structures of the 14 selected fragments

5.6 Summary

A small library of fragment compounds has been selected for use in exploratory research into NMR screening. Such fragments are expected to interact with the SH2 domain of *v*-Src SH2 on the basis of favourable binding free energy predictions and homology to known binders. The success of these approaches will be verified through

in vitro analysis, described in the next chapter. Results from molecular docking appeared to be more robust than those from chemical similarity searching, perhaps because the former technique was more discerning. Indeed, it is the binding site that determines the structure and chemistry of the ligand and not the other way round. The hydrophilic nature of the pY pocket imposes relatively narrow constraints on the structure of acceptable ligands compared to the largely hydrophobic surface of the pY+3 pocket. In particular, the hydrogen bonding pattern that dominates SH2 domain interactions requires that strict distance and angular criteria between certain atom types are adhered to. The validity of molecular docking calculations with *v*-Src SH2 is supported by the backbone dynamics and pK_a data presented in Chapter 4, which indicated that the phosphotyrosine pocket is well defined by local secondary structure and an important hydrogen bonding network that controls the orientation of ArgβB5 (R32). The structural changes required to interact with a non-standard ligand would therefore present an insurmountable energetic barrier to binding, and explains the extreme specificity towards phosphotyrosine. In this case, a search for compounds that make the same set of interactions as the natural ligand is perhaps the most direct route to obtaining a novel binding fragment. There is some plasticity in the SH2 domain binding site that was not modelled by FLEXX. It would be interesting to observe which additional compounds are identified through use of the more advanced algorithm, FLEXE, which permits fluctuation of nearby side chains and the BC loop.

Chapter 6

In vitro fragment screening

The compounds identified through computational screening procedures represent a small fragment library for use in the assessment of NMR screening strategies. This chapter describes the exploration of one such approach – a ^{19}F-based, competition-style NMR screening experiment known as *fluorine chemical shift anisotropy and exchange for screening* (FAXS)[93]. As described in Section 1.2.3, FAXS relies on detecting changes in the observed transverse relaxation rate ($R_{2,obs}$) of a reporter ligand (Spy molecule), which is dependent on the presence or absence of competing ligand. Reversible interaction with the receptor causes nuclei in the reporter ligand to undergo rapid relaxation and leads to a controllable reduction of peak intensity in the NMR spectrum. If this interaction is inhibited by competition with another ligand, $R_{2,obs}$ is closer to that of free reporter ligand and yields a larger peak intensity. The presence of a ligand with a given screening mixture is therefore proved by an increase in the reporter ligand NMR signal.

In principle, the competition-style, R_2-edited screening approach exemplified by FAXS could be applied to any 'NMR-active'* nucleus in the reporter ligand. Indeed the progenital versions of FAXS employed proton spectra[90], however in these experiments, there is strong potential for resonance overlap between reporter molecules and other system components such as test compounds, buffers, solvents, detergents, and the receptor. This presents problems for the automated analysis of spectra when screening mixtures. Whilst ^{13}C- or ^{15}N-labelled reporter molecules could be employed instead, their inherent low signal strength coupled to the high cost of incorporating such nuclei into the Spy molecule presents an equally unfavourable set of problems. ^{19}F is a very attractive alternative due to its high sensitivity (83 % that of proton), 100 % natural abundance, large chemical shift range, and small chance of interfering spectral overlap. Moreover, fluorination of ligands is chem-

*Nuclei that have a spin quantum number greater than zero display nuclear magnetic resonance, and are loosely termed 'NMR-active'.

ically straightforward and has been increasingly applied during lead-optimisation processes[93].

FAXS was optimised for the SH2 model system and used to screen the fragment library. Where relevant, fragment binding was also investigated using isothermal titration calorimetry and $^{1}H^{15}N$-HSQC titration experiments. The main experimental section of this chapter is preceded by a theoretical discussion of FAXS in terms of factors determining $R_{2,obs}$, its measurement using the CPMG pulse sequence, and the competition binding equilibrium that governs the fraction of Spy molecule bound to the receptor during the experiment. These subjects will be discussed in turn, using simulations to illustrate the relevant equations. Proper understanding of FAXS theory also provides a rational basis for selection of useful Spy molecules, a subject that is also discussed below.

6.1 Theoretical considerations

6.1.1 Transverse relaxation rates

R_2 describes the rate at which a population of spin vectors loses phase coherence. This is caused by dynamic changes in the local magnetic field that cause nuclear spins to precess at slightly different frequencies, which eventually leads to zero net magnetisation in the xy-plane (see Figure 1.10 on page 41). Loss of phase coherence also occurs when the nuclear spin undergoes chemical exchange processes, such as reversible binding to another molecule. This may be a direct effect, or it may be indirectly caused by magnetic coupling to nearby affected nuclei. Where the longitudinal (R_1) relaxation rate of a coupled spin is relatively fast, there will be an additional contribution to R_2.

The rotational correlation time (τ_c) of the molecule is also a strong determinant of R_2. This random motion occurs at a spread of frequencies where τ_c represents the arithmetic mean time taken to rotate through one radian. The frequency distribution of the fluctuating magnetic field associated with molecular rotation can be described using the spectral density function, $J(\omega)$, discussed in Chapter 4. This has the general form

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 {\tau_c}^2}. \tag{6.1}$$

For slowly-tumbling molecules there is a higher probability that there will be a magnetic field fluctuating at the Larmor frequency than for rapidly-tumbling molecules. Therefore R_2 is proportional to τ_c, and also to molecular weight.

6.1.2 Measurement of $R_{2,obs}$

In practice, inhomogeneities in the external magnetic field also contribute to $R_{2,obs}$ since molecules diffuse through regions of altered field strength, or experience a time-dependent fluctuation in the field. Observed transverse relaxation rates ($R_{2,obs}$) are therefore represent summed contributions from genuine molecular mechanisms (R_2) and artefactual field inhomogeneities ($R_{2(\Delta B_0)}$):

$$R_{2,obs} = R_2 + R_{2(\Delta B_0)} \tag{6.2}$$

Since the relaxation mechanisms that contribute to longitudinal relaxation also lead to the loss of phase coherence and that return of the bulk magnetisation vector to the z-axis abolishes net magnetisation within the xy-plane, R_2 can never be slower than R_1. For rapidly-tumbling molecules, such as fragments, R_2 and R_1 tend to be equal however as molecular weight and dynamic complexity increases, additional contributions to R_2 are apparent.

Measurement of the natural R_2 can only be achieved by removing the artefactual contribution from magnetic field inhomogeneity. In practice this is performed using 'spin-echo'[329] pulse sequences (Figure 6.1). It is assumed that the external magnetic field consists of a 3D array of microscopic regions within which the magnetic field is perfectly homogenous and stable. Spin vectors within such regions therefore precess at identical frequencies, and are termed isochromats. The spin-echo experiment begins with a 90°_x pulse that rotates the equilibrium net magnetisation vector from the $+z$-axis to the $+y$-axis. During a subsequent delay period (τ), the isochromats begin to lose phase coherence with each other since they experience slightly different field strengths. A 180°_x pulse rotates the vectors towards the -y-axis, and during a second identical delay, the isochromats begin to regain phase coherence until they align perfectly along the -y-axis. Whilst this sequence cancels out the effects of field inhomogeneity, natural R_2 relaxation processes will still cause individual spin vectors to refocus imperfectly, and lead to a reduced signal intensity for the echo. R_2 can be extracted by arraying the τ delay and fitting the change in intensity to an exponential decay function[330].

The basic spin-echo method assumes that field inhomogeneity is static and that molecules do not diffuse into regions of altered field. In reality, neither premise is valid and as the delay period is lengthened, R_2 is increasingly overestimated. Carr and Purcell[331] reduced the influence of these problems by keeping τ short and repeating the τ–180°_x–τ sequence many times (Figure 6.1). However, if the 180°_x pulse length is imperfect then refocusing errors will be cumulative, again contributing to $R_{2,obs}$. Meiboom and Gill[73] circumvented this difficulty by applying the 180° pulses

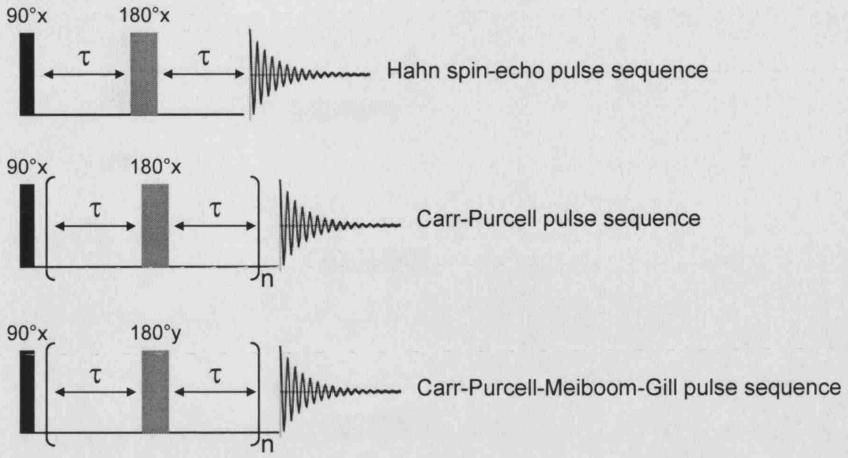


Figure 6.1 Schematic diagrams of the Hahn spin-echo, Carr-Purcell, and Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences. Radiofrequency pulses are indicated by filled rectangles.

along the y-axis instead, which causes refocusing to occur along the $+y$-axis for each echo. Errors in the 180°_y pulse are small and constant for every odd-numbered echo, but cancel out for the even-numbered echo. The CPMG pulse sequence (Figure 6.1) can be used to measure R_2 by fitting change in intensity for even-numbered echoes against the total delay period ($n2\tau$).

The CPMG sequence provides a method for selectively editing resonances corresponding to rapidly-relaxing nuclei out of the spectrum. The $n2\tau$ period can be optimised such that signals from macromolecules or molecular complexes (e.g. proteins and membranes) can be suppressed whilst those of smaller compounds remain visible in the spectrum. For NMR screening purposes, the CPMG experiment is used to detect the significant increase in R_2 displayed by a ligand through interaction with a receptor. It is this approach that is employed by FAXS to monitor the binding behaviour of a reporter ligand in the presence of competitive ligands.

The signal intensity corresponding to the ^{19}F nucleus (or nuclei) in the reporter ligand is dependent on physical quantities (free and bound transverse relaxation (R_2) and translational diffusion (D) rates), molecular interactions (bound state lifetime ($1/k_{-1}$), chemical shift changes ($\omega_f - \omega_b$), fractional ligand saturation ([PL]/L_{tot})) and experimental NMR parameters ($n2\tau$, field inhomogeneity (G))[93]. Essentially, the available signal at the end of the CPMG pulse sequence is modulated by $R_{2,obs}$ and any contributions from diffusion within the non-homogenous magnetic field. These contributions are described mathematically in Equations 6.3–6.5, where I_0 represents the available signal immediately after the 90°_x pulse, at which time no relaxation has occurred, and γ_F is the gyromagnetic ratio of ^{19}F.

$$I = I_0 e^{-\gamma_F^2 G^2 D_{obs}(n2\tau)\tau^2/3} e^{-n2\tau R_{2,obs}} \quad (6.3)$$

$$\text{where} \quad D_{obs} = \frac{[PL]}{L_{tot}} D_b + \left(1 - \frac{[PL]}{L_{tot}}\right) D_f \quad (6.4)$$

$$\text{and} \quad R_{2,obs} = \frac{[PL]}{L_{tot}} R_{2,b} + \left(1 - \frac{[PL]}{L_{tot}}\right) R_{2,f} + \frac{[PL]}{L_{tot}} \left(1 - \frac{[PL]}{L_{tot}}\right)^2 \frac{4\pi^2(\omega_f - \omega_b)^2}{k_{-1}} \quad (6.5)$$

The diffusion rate (D_{obs}) decreases linearly as the fraction of bound ligand increases. If the contribution of the third term in Equation 6.5 is negligible, then $R_{2,obs}$ is also linearly correlated to $[PL]/L_{tot}$. This occurs when the chemical shift changes experienced on binding are very small, or when the residence time ($\tau_{res} = 1/k_{-1}$) is very short. The above equations are only valid for molecular interactions in the fast to intermediate exchange timescales (Dr Claudio Dalvit, personal communication).

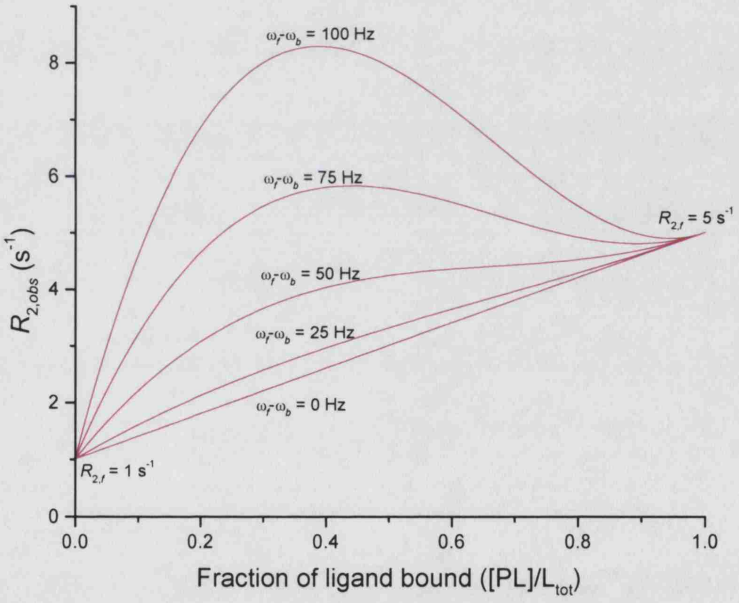


Figure 6.2 Simulated relationship between the fraction of ligand bound ($[PL]/L_{tot}$) and observed R_2 relaxation rates according to Equation 6.5. The free and bound state R_2 values were 1 and 5 s^{-1}, respectively, k was 5×10^4 s^{-1}. Curves are simulated for several values of $\omega_f - \omega_b$, from 0 to 100 Hz, as labelled on the graph.

Changes in $R_{2,obs}$ with fractional ligand saturation are simulated in Figure 6.2 using a variety of contributions from chemical exchange. Clearly exchange processes can have a significant effect on $R_{2,obs}$, especially at intermediate levels of ligand saturation. Simulations were also performed to explore the effect of experimental

parameters on the signal intensity at the end of the CPMG pulse sequence (Figure 6.3). The difference in R_2 relaxation rates for the free and bound ligand is a strong determinant of the available signal intensity. Where this difference is large, only small amounts of protein are required to induce rapid relaxation of signal for the reporter ligand. Consequently, FAXS becomes more efficient when used with larger proteins. Finally, the effect of increasing the total CPMG delay period ($n2\tau$) is to shift the curve to lower intensity (i.e. the initial intensity is lower, however less protein is required to achieve complete suppression of signal). These simulations suggest that FAXS is most efficient when differences in R_2 are inherently large, especially when enhanced by chemical exchange contributions. The $n2\tau$ delay can be optimised to generate the desired signal intensity at the level of ligand saturation employed.

6.1.3 Binding equilibria

The equations relevant to FAXS discussed above include a term ($[PL]/L_{tot}$) that reflects the fraction of ligand molecules that are bound during the experiment. This quantity will depend on the relative concentrations of protein and ligand, as well as the affinity of interaction. The purpose of this section is to describe the general binding equilibria expected for a protein–ligand interaction, and is drawn from several published sources[†].

The association of a protein and a ligand can be described as a dynamic equilibrium ($[P] + [L] \rightleftharpoons [PL]$) in which the off rate (k_{off}) is inversely proportional to the mean lifetime of the interaction and the on rate (k_{on}) is a measure of the probability of successful interaction between the two species. The binding affinity is described by a dissociation constant K_d, which is equal to both $[P][L]/[PL]$ and k_{off}/k_{on}. The fractional occupation of protein binding sites (X_{Pb}) is a hyperbolic function of [L] such that when $[L] = K_d$, the receptor is 50 % saturated (Equation 6.6). The concentration of the protein-ligand complex ([PL]) can also be expressed in terms of known experimental concentrations P_{tot} and L_{tot}, which represent the total protein and ligand concentrations respectively (Equation 6.7).

$$X_{Pb} = \frac{[PL]}{[P] + [PL]} = \frac{[L]}{[L] + K_d} \tag{6.6}$$

$$[PL] = \left(P_{tot} + L_{tot} + K_d - \sqrt{(P_{tot} + L_{tot} + K_d)^2 - 4P_{tot}L_{tot}}\right)/2 \tag{6.7}$$

$$\text{where } P_{tot} = [PL] + [P]$$

$$\text{and } L_{tot} = [PL] + [L]$$

[†]Meyer & Peters[68], Lepre *et al.*[69], Price & Dwek[332], and Stryer[293].

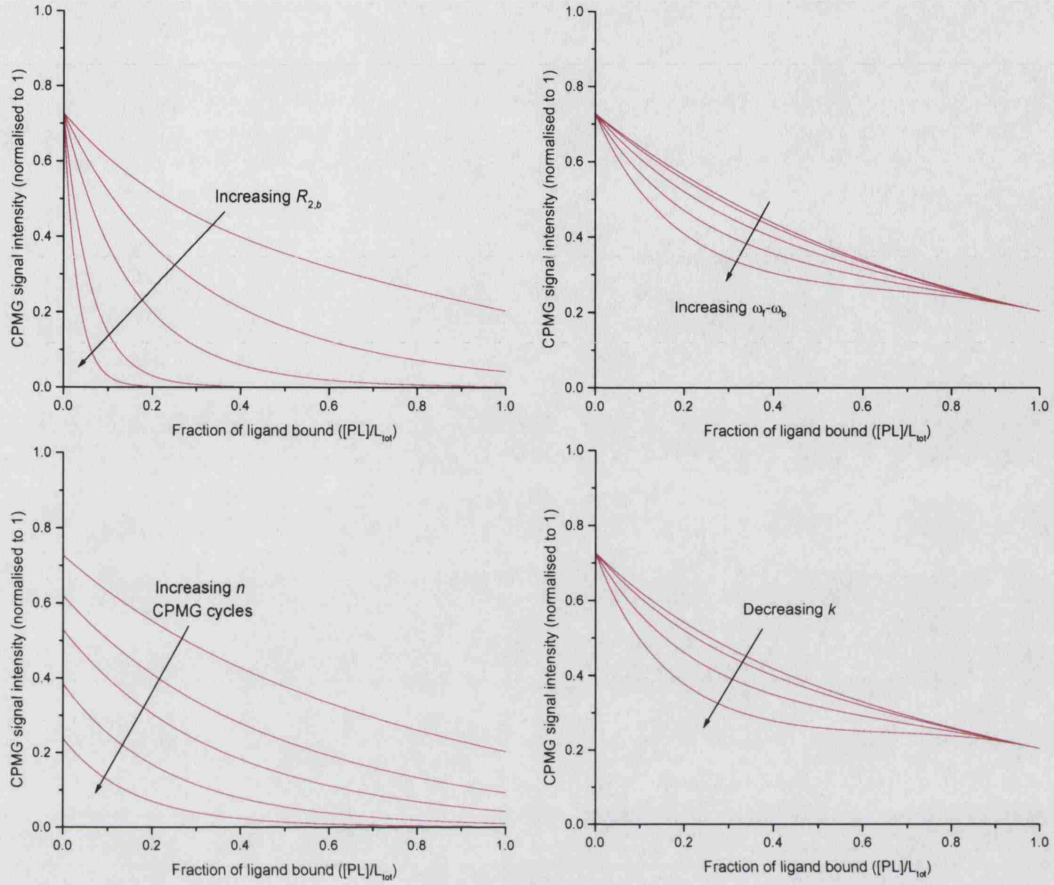


Figure 6.3 Simulations of FAXS signal intensity changes with fraction of bound ligand, according to Equations 6.3 and 6.5. Perfect field homogeneity was assumed throughout, thus the first exponential term of Equation 6.3 was ignored in each simulation. Unless otherwise stated, default simulation parameters were as follows: $R_{2,f} = 1\ \mathrm{s}^{-1}$, $R_{2,b} = 5\ \mathrm{s}^{-1}$, $\omega_f - \omega_b = 25$ Hz, $k = 5 \times 10^4\ \mathrm{s}^{-1}$, CPMG loop counter $(n) = 8$, and delay $(\tau) = 20$ ms. *Top left* | The effect of altering the difference between free and bound R_2 values was explored by fixing $R_{2,f}$ at $1\ \mathrm{s}^{-1}$ and varying $R_{2,b}$ (5,10,20,50,100 s^{-1}). *Top right* | Increasing chemical exchange contributions were investigated by varying $\omega_f - \omega_b$ (0,25,50,75,100 Hz). *Bottom left* | Increasing the CPMG loop counter n (8,12,16,24,36) leads to a reduction in signal intensity. *Bottom right* | Decreasing values of k (i.e. increasing affinity, $1 \times 10^5, 5 \times 10^4, 2 \times 10^4, 1 \times 10^4\ \mathrm{s}^{-1}$) leads to a reduction in signal intensity.

Figure 6.4 shows a simulated titration of ligand into protein using Equation 6.7 and three alternative values of K_d. It is clear that the concentration of ligand required to saturate protein binding sites is proportional to K_d. The fraction of bound ligand ($X_\mathrm{Lb} = [\mathrm{PL}]/\mathrm{L_{tot}}$) occurs within the range $0 \rightarrow \mathrm{P_{tot}}/\mathrm{L_{tot}}$. The change in X_Lb with $\mathrm{L_{tot}}$ is also simulated in Figure 6.4. Clearly, the fraction of bound ligand is maximal for small values of $\mathrm{L_{tot}}$, however saturation is also reached for very high levels of $\mathrm{L_{tot}}$.

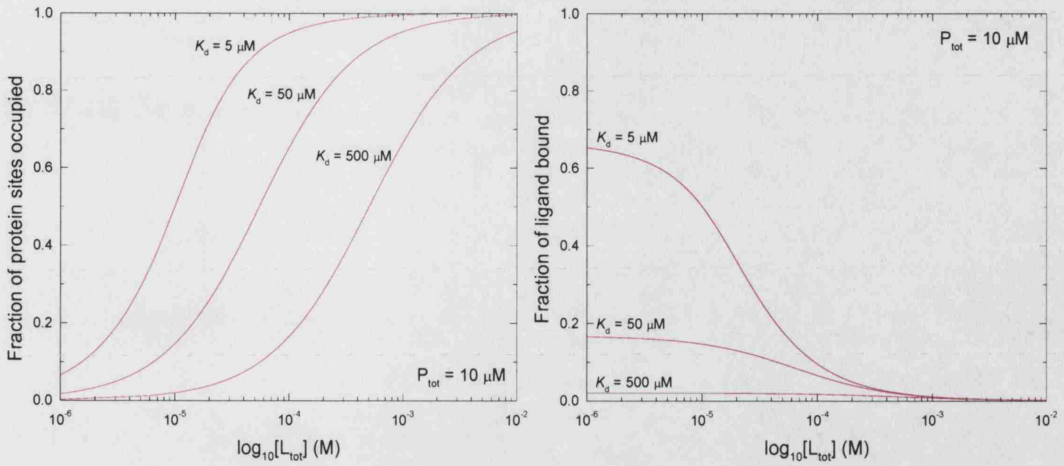


Figure 6.4 Simulated protein-ligand binding equilibria for three values of K_{d}. Plots indicate the effect of increasing ligand concentration $\mathrm{L_{tot}}$ on the fractional saturation of protein sites (*left*) and fraction of ligand bound (*right*). In both simulations, protein concentration was fixed at 10 μM.

6.1.4 Competition binding equilibria

Introduction of an inhibitor, I, that competes with the ligand for the same protein binding site leads to a reduction in [PL], dependent on their relative dissociation constants, K_{i} and K_{d} and concentrations (Equation 6.11). If K_{i} is known, then it is possible to estimate the K_{d} of L through titration experiments:

$$\mathrm{I_{tot}} = [\mathrm{PI}] + [\mathrm{I}] \tag{6.8}$$

$$\mathrm{L_{tot}} = [\mathrm{PL}] + [\mathrm{L}] \tag{6.9}$$

$$\mathrm{P_{tot}} = [\mathrm{PL}] + [\mathrm{PI}] + [\mathrm{P}] \tag{6.10}$$

$$[\mathrm{PL}]/\mathrm{L_{tot}} = \frac{[\mathrm{L}]}{[\mathrm{L}] + K_{\mathrm{d}}^{*}} \tag{6.11}$$

$$\text{where } K_{\mathrm{d}}^{*} = K_{\mathrm{d}}\left(1 + \frac{[\mathrm{I}]}{K_{\mathrm{i}}}\right) \tag{6.12}$$

The fractions of protein sites bound to L and I (X_{PL} and X_{PI}, respectively) can be expressed in terms of known experimental parameters $\mathrm{P_{tot}}$, $\mathrm{L_{tot}}$, and $\mathrm{I_{tot}}$ as well as K_{d} and K_{i}:

$$X_{\mathrm{PL}} = \frac{[\mathrm{PL}]}{\mathrm{P_{tot}}} = \frac{(\mathrm{L_{tot}}/\mathrm{P_{tot}})X_{\mathrm{P}}}{K_{\mathrm{d}}/\mathrm{P_{tot}} + X_{\mathrm{P}}} \tag{6.13}$$

$$X_{\mathrm{PI}} = \frac{[\mathrm{PI}]}{\mathrm{P_{tot}}} = \frac{(\mathrm{I_{tot}}/\mathrm{P_{tot}})X_{\mathrm{P}}}{K_{\mathrm{i}}/\mathrm{P_{tot}} + X_{\mathrm{P}}} \tag{6.14}$$

$$\text{where } X_{\mathrm{P}} = [\mathrm{P}]/\mathrm{P}_{\mathrm{tot}} \tag{6.15}$$

By fixing $X_{\mathrm{P}} + X_{\mathrm{PL}} + X_{\mathrm{PI}} = 1.0$, the fraction of free protein, X_{P}, can be obtained:

$$X_{\mathrm{P}} = \frac{2\sqrt{\alpha^2 - 3\beta}\cos(\Theta/3) - \alpha}{3} \tag{6.16}$$

$$\text{where } \Theta = \cos^{-1}\left\{\frac{-2\alpha^3 + 9\alpha\beta - 27\gamma}{2\sqrt{(\alpha^2 - 3\beta)^3}}\right\} \tag{6.17}$$

$$\alpha = (1/\mathrm{P}_{\mathrm{tot}})(K_{\mathrm{i}} + K_{\mathrm{d}} + \mathrm{I}_{\mathrm{tot}} + \mathrm{L}_{\mathrm{tot}} - \mathrm{P}_{\mathrm{tot}}) \tag{6.18}$$

$$\beta = (1/\mathrm{P}_{\mathrm{tot}})^2\{K_{\mathrm{i}}(\mathrm{L}_{\mathrm{tot}} - \mathrm{P}_{\mathrm{tot}}) + K_{\mathrm{d}}(\mathrm{I}_{\mathrm{tot}} - \mathrm{P}_{\mathrm{tot}}) + K_{\mathrm{i}}K_{\mathrm{d}}\} \tag{6.19}$$

$$\gamma = -K_{\mathrm{i}}K_{\mathrm{d}}/{\mathrm{P}_{\mathrm{tot}}}^2 \tag{6.20}$$

Finally, substitution of X_{P} into Equations 6.13 and 6.14 produces [PL] and [PI]. The competitive displacement of reporter ligand by the inhibitor is simulated in Figure 6.5 using a range of inhibitor affinities. As [I] increases, the concentration of protein binding sites available for interaction with L decreases, causing L to be displaced into solution. Competition is greater for higher affinities of inhibitor.

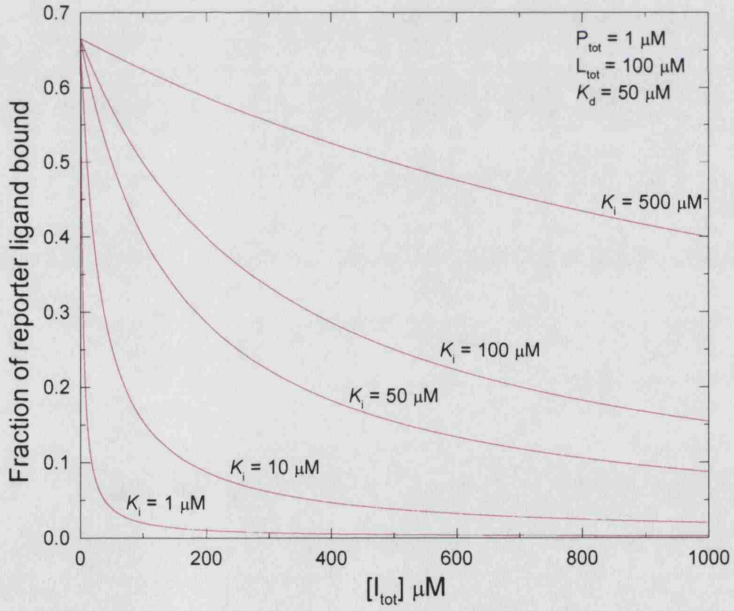


Figure 6.5 Simulated competitive displacement of ligand L by inhibitor I according to several values of K_{i}. During this simulation, $\mathrm{P}_{\mathrm{tot}}$, $\mathrm{L}_{\mathrm{tot}}$, and K_{d} were fixed at values shown in the top right-hand corner of the graph.

6.1.5 Rational selection of reporter compounds for FAXS

Successful use of FAXS in fragment screening relies on careful selection of the Spy molecule. The constraints on this compound are two-fold: it must bind to the receptor within an particular affinity range (300 μM $\geq K_d \geq$ 0.1 μM) and it must be fluorinated. If the K_d is too low then only a small proportion of ligand will bind and undergo rapid relaxation; the remainder will relax slowly and thus $R_{2,obs}$ will remain closer the R_2 of the free ligand. Under this regime, it is difficult to detect competition between the Spy molecule and other ligands.

The fluorine labelling strategy also affects the overall sensitivity of FAXS. As the above simulations show, chemical exchange can have a profound and useful effect on $R_{2,obs}$ at partial levels of reporter ligand saturation. Positioning of the fluorine nucleus at a site where it experiences a significant perturbation in chemical shift enhances the contribution from chemical exchange. This increases the net relaxation rate of the reporter ligand and thus has a favourable effect on experimental sensitivity and protein requirements.

Dalvit and co-workers suggest labelling reporter compounds with either a CF or CF_3 group[93]. Trifluoromethyl groups generate a stronger signal than mono-fluorinated compounds, which allows use of lower compound concentrations in screening. However, the rapid rotation of fluorine atoms around the C_3 axis leads to larger intrinsic R_2 rates compared to CF groups, which reduces the difference in transverse relaxation rates between the free and bound states, and therefore has a negative impact on experimental sensitivity. Addition of a bulky CF_3 group to a ligand may also provide unfavourable steric hindrance to binding.

Reporter ligands containing a CF group are more suited to the FAXS approach as the difference in relaxation rates between the free and bound states are often quite large. The small size of this moiety means that insertion at protein-ligand interfaces is more likely to be tolerated. Moreover, CF groups can act as hydrogen bond acceptors. Both these factors will favour a significant contribution from chemical exchange to $R_{2,obs}$, and increase experimental sensitivity. Dalvit *et al.* also include a non-binding, fluorinated 'Control' molecule in competition experiments. This is employed as an internal standard to enable accurate integration of the Spy molecule peak.

6.2 Implementation of FAXS

The FAXS screening experiment was successfully implemented on the 400 MHz Bruker spectrometer at Celltech (Cambridge, UK) using the *v*-Src SH2 domain and the library of fragment compounds selected from computational modelling. *In silico*

screening highlighted three ^{19}F-containing phosphotyrosine mimetic compounds that could be useful as Spy molecules in FAXS, two of which were obtained (Figure 6.6). A third compound, fluorophosphate (FPO_3), was also used as a low-affinity reporter ligand. In each case, the phosphate moiety ensured that the aqueous solubility of Spy molecules was very high. Hexafluoroacetone (CF_3–C(=O)–CF_3) was selected to be the Control molecule as it displayed no affinity for *v*-Src SH2, is highly soluble in aqueous buffers, and generates a very intense NMR signal.

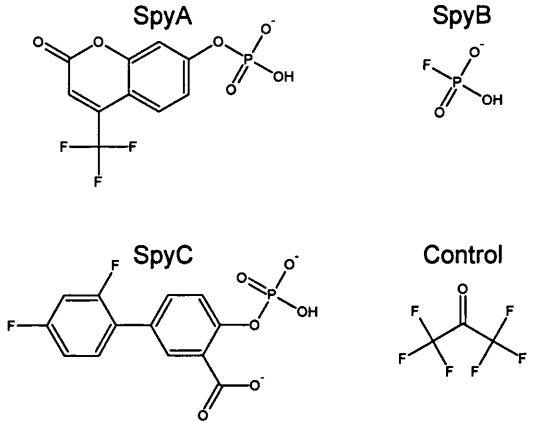


Figure 6.6 Spy and Control molecule structures

Before the computer-derived library of fragment compounds could be screened, experimental parameters had to be optimised. The original FAXS publication employed a monofluorinated Spy molecule with a dissociation constant of 41 μM and monitored binding to human serum albumin (HSA, 66 kDa) on a 600 MHz (14.7 T) spectrometer[93]. The system under investigation in this project presents a less favourable situation in several respects. Firstly, the SH2 domain is five times smaller than HSA, and therefore the bound Spy molecule relaxation rate is likely to be significantly lower than for the equivalent reporter ligand interacting with HSA. Secondly, the K_d of the selected compounds and reporter ligands are likely to exceed 100 μM resulting in short ligand residence times in the binding site. Consequently, protein concentration will have to be raised in order to increase the number of binding events necessary to induce relaxation of the Spy molecule. Thirdly, the lower magnetic field strength employed (9.4 T) decreases the overall sensitivity and has a negative impact on the contribution of chemical exchange to $R_{2,obs}$[‡].

The FAXS experiment was optimised by varying CPMG pulse sequence parameters and sample concentrations until the signal from Spy molecule was significantly reduced in the presence of SH2 domain (see below). A known binder, phosphotyrosine, was used to compete with the Spy molecule, and effect a recovery of

[‡]Chemical shift differences between the free and bound states of the Spy molecule are proportional to field strength.

the NMR signal. Thermodynamic binding data describing the interaction between SpyA, SpyB, pY and the SH2 domain provided an indication of the (competitive) binding equilibria in operation.

6.2.1 Preparation of materials

Fragment compounds were solubilised to a concentration of 5–50 mM in NMR buffer[§]. A small volume of dimethylsulphoxide (DMSO) was used to dissolve many of the solids, however this was diluted to 0.5–2 % [v/v] in the stock solutions, and to even lower levels within each NMR sample. Previous studies have shown that even at DMSO concentrations of 5 % [v/v], the backbone chemical shifts of *v*-Src SH2 are largely unaffected (personal observation, data not shown). Compounds were stored as necessary at either room temperature, 4 °C or -20 °C. Unlabelled *v*-Src SH2 polypeptide was prepared as described in Section 2.2.1, and stored in NMR buffer at a concentration of 0.6 mM.

6.2.2 Thermodynamic profile of fragment binding

ITC was used to analyse the thermodynamics of binding for phosphotyrosine, SpyA and SpyB. As will be discussed in Section 6.4.3 below, unresolvable complexities in the NMR spectrum of SpyC prevented its use in FAXS, and so it was not analysed by ITC. A direct titration of phosphotyrosine into *v*-Src SH2 did not generate useful calorimetric data as its affinity was at the lower limit of detection ($c \approx 0.2$[¶]). A previous attempt by Bradshaw *et al.*[106] utilised 1.6 mM *v*-Src SH2 domain and 30 mM pY ($c = 4.8$) and resulted in a K_d value of 333 μM. In the buffer conditions employed for these studies, significant precipitation of *v*-Src SH2 occurred at concentrations in excess of 800 μM.

Thermodynamic binding parameters for low-affinity compounds such as pY are accessible using a competition-style ITC approach[229, 230]. The presence of a low-affinity competitor affects the observed binding thermodynamics for a high-affinity ligand, from which the dissociation constant and binding enthalpy of the competitor can be extracted. A full mathematical treatment of this approach was provided by Zhang & Zhang[230]. This strategy was used to calculate thermodynamic data for the three low-affinity fragment ligands, using the high-affinity phosphopeptide (PQpYEEIPI) as the coupled ligand.

The thermodynamic profile of the SH2–peptide interaction was first characterised in the absence of competitive ligand using standard ITC methodology: Serial injec-

[§]20 mM MES, pH 6.00, 50 mM NaCl, 1 mM DTT, 10 % [v/v] D_2O

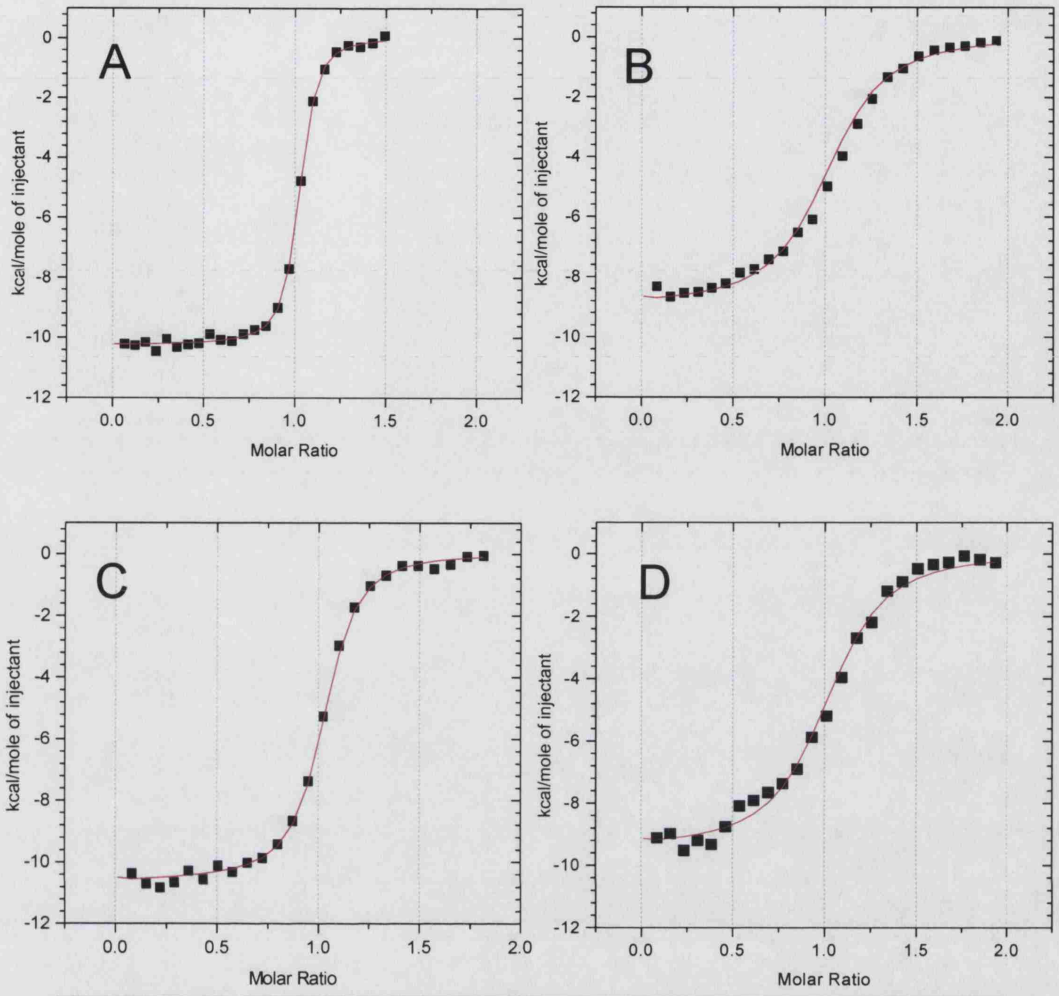
[¶]For a description of the *c* value see Section 2.4.

tions ($n = 24$) of peptide (0.24 mM) into buffer, and subsequently into a solution containing *v*-Src SH2 (35 μM, $c = 500$) yielded thermodynamic values that were very similar to those described previously (Figure 6.7)[106, 117]. The two experiments were then repeated in the presence phosphotyrosine (initial cell concentration = 2 mM), which competes for SH2 domain binding sites and therefore affects the apparent binding thermodynamics of the peptide. The resultant titration data (Figure 6.7) were fitted in *ORIGIN* using a script describing competition equilibria provided by Dr Bent W. Sigurskjold[229,‖].

Identical ITC experiments were performed using SpyA or SpyB as low-affinity, competitive ligands instead of pY (Figure 6.7). In each case, fragment binding energy is dominated by a favourable entropic term that probably corresponds to the release of water molecules solvating ligand and protein, although this is offset by loss of configurational freedom in protein/ligand bonds. As might be expected, the most favourable entropy is observed for SpyB (fluorophosphate), which is devoid of rotatable bonds and is unlikely to lead to trapping of solvent molecules, nor cause significant restriction of protein side chains. Both pY and SpyA indicate a favourable enthalpy change suggesting that solvent–protein and solvent–fragment interactions have a higher enthalpy than protein–fragment and solvent–solvent interactions. The binding thermodynamics of SpyA are close to those observed for phosphotyrosine, indicating that the way in which these two compounds interact with the *v*-Src SH2 is energetically similar. The common structural feature of each ligand is an aromatic phosphate group. Equivalence in binding thermodynamics suggests that it is this moiety that provides the important contacts rather than the amino/acid part of phosphotyrosine, or trifluoromethyl part of SpyA.

SpyB indicates a significantly different thermodynamic profile to the other fragments. The positive enthalpy change on binding could indicate either a strenghening of interactions with solvent, or a weakening of interactions with the SH2 domain, or both. Affinity of *v*-Src SH2 for fluorophosphate is almost an order of magnitude lower than for phosphotyrosine. This loss is probably due to removal of the π-interactions between R12/K60 and the aromatic ring and/or reduced hydrogen bonding capability through exchange of an oxygen for a fluorine. The reduced affinity of SpyB makes it a useful reporter ligand for screening very weak-binding library compounds that would not successfully compete with SpyA.

‖Department of Biochemistry, August Krogh Institute, University of Copenhagen, Denmark



Ligand	K_d (μM)	ΔH_b° (KJ.mol^{-1})	TΔS (KJ.mol^{-1})	ΔG_b° (KJ.mol^{-1})
PQpYEEIPI	0.10 ± 0.01	-34.8 ± 0.15	5.13	-39.93
pY	175 ± 27	-3.30 ± 0.49	18.13	-21.43
SpyA	177 ± 29	-5.53 ± 0.53	15.87	-21.40
SpyB	1627 ± 205	+13.86 ± 0.38	29.77	-15.91

Figure 6.7 ITC-derived binding thermodynamics for high-affinity peptide, and three fragment ligands. Data for the peptide were obtained using standard ITC methodology whilst competition experiments were employed for the analysis of low-affinity fragments. Fitted binding isotherms are presented for peptide (A), phosphotyrosine (B), SpyA (C) and SpyB (D) and full thermodynamic data presented in the table below.

6.3 Optimisation of FAXS for fragment screening

6.3.1 Optimisation of CPMG pulse sequence

The pulse parameters and delays within the CPMG experiment require optimisation for each molecular system. The 90°_x pulse width that generated maximal signal for SpyA was 10.5 μs at power attenuation of 0 dB (high power). The 180°_y pulse width

was obtained by doubling this value. The vast majority of NMR experiments are performed using multiple 'scans' in order to improve the signal to noise ratio. It is imperative to allow sufficient delay between scans in order that the population of spins regains the initial equilibrium state, which occurs at the exponential rate R_1. This relaxation rate can be determined using the inversion-recovery sequence[333]. In this scheme, the delay period (τ_{null}) that gives rise to zero signal at the point of acquisition can be used to estimate T_1 as follows:

$$T_1 = \tau_{\text{null}} / \ln 2 \tag{6.21}$$

The T_1 for the trifluoromethyl group of SpyA was 0.72 s, which imposes a 3.6 s delay between the final pulse of one scan and the first of the next (this can include acquisition time)**.

A proton CPMG pulse sequence was modified for ^{19}F-NMR, inserting the appropriate pulse widths and delays (Figure 6.8). Spectral sweep width was set throughout to 25 ppm (9416 Hz), with the offset dependent on the choice of spy molecule (e.g. -74 ppm for SpyA). Each experiment started with 4 dummy scans in which no data was collected, and, depending on resolution required, between 16 and 256 scans were performed, collecting 32,000 data points in the acquisition period††. Dalvit *et al.*[93] employed a τ delay of 20 ms, which was effective in these studies.

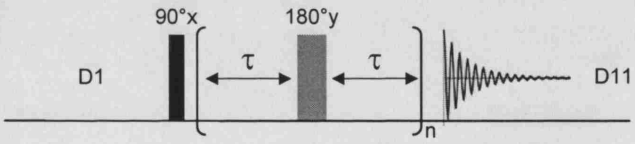


Figure 6.8 CPMG pulse sequence with each pulse and delay shown, not to scale. Relaxation delay D1 = 2 s; 90°_x = 10.5 μs; τ delay = 20 ms; 180°_y = 21 μs; acquisition time = 1.74 s; file writing delay D11 = 30ms. Total scan time = 4.1 s

The relaxation behaviour of free SpyA and Control molecules was investigated by altering the total CPMG delay ($n2\tau$, where $n = 2, 10, 20, 50$). The signal intensity of both molecules decreased as the n parameter was increased (data not shown), indicating that the CPMG pulse sequence was functioning correctly (signal should decay exponentially with n). Complete relaxation of SpyA signal required a total delay of 2.0 seconds ($n = 50$).

**Empirically, it is found that after a period of 5*T_1, 99.3 % of transverse magnetisation has decayed to zero[330]

††Due to phase cycling for the different scans, no. scans ÷ 8 must yield an integer value.

6.3.2 Observation of SpyA – SH2 interaction

It is expected that addition of SH2 domain to a sample of SpyA and Control would lead to a selective decrease in reporter ligand signal intensity. CPMG experiments were performed using several samples of SpyA (200 μM) and Control (100 μM), containing *v*-Src SH2 at concentrations between 0 and 200 μM. The CPMG delay period was also varied for each sample ($n = 2, 4, 6, 8$) in order to establish the optimal experimental parameters for future screening. The signal intensity of SpyA decreased as the concentration of receptor molecule increased, which strongly suggests an interaction between the two molecules (Figure 6.9). Conversely, peaks integrals measured for the Control molecule were not sensitive to the addition of SH2 domain, indicating that its R_2 is unaffected, and hence does not bind. NMR spectra corresponding to $n = 8$ are displayed in Figure 6.9 where complete relaxation of the SpyA signal occurs at equimolar SH2 domain and reporter ligand concentrations.

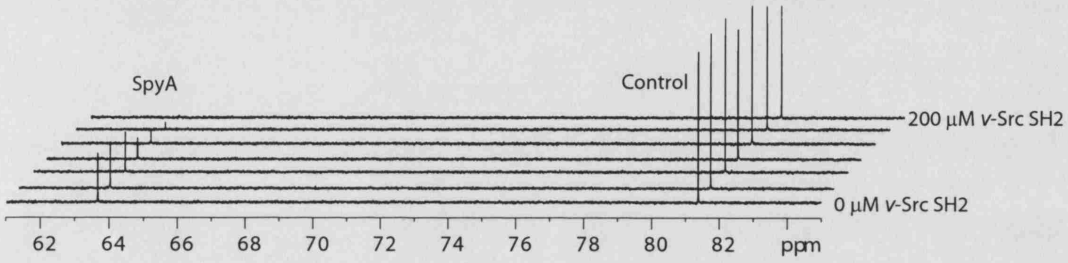


Figure 6.9 Effect of SH2 domain concentration on the CPMG spectrum of SpyA (200 μM) and Control (100 μM). The chemical shifts of the reporter molecules are -63.7 (SpyA) and -81.3 ppm (Control). The stacked spectra correspond to SH2 domain concentrations of 0, 1, 10, 50, 100, 150, and 200 μM. All CPMG spectra were recorded using $n2\tau = 8 \times 2 \times 0.02 = 0.32$ s.

6.3.3 Observation of SpyA – phosphotyrosine competition

The previous section demonstrated that SpyA interacts with *v*-Src SH2 and indicated potential concentration regimes for use in FAXS screening. The effects of a competitive ligand (pY) was subsequently explored in several samples containing equimolar SpyA and SH2 domain (200 μM), Control (50 μM), and a range of pY concentrations (0.1–13400 μM). In CPMG spectra recorded for each sample, the intensity of the SpyA molecule resonance was found to be proportional to the concentration of pY, indicating that both ligands compete for the same binding site (Figure 6.10).

In theory, the data shown in Figure 6.10 could be used to obtain information regarding the dissociation constant of SpyA or pY, depending on which is classed as the known variable. Unfortunately, this quantitative analysis is complicated by the

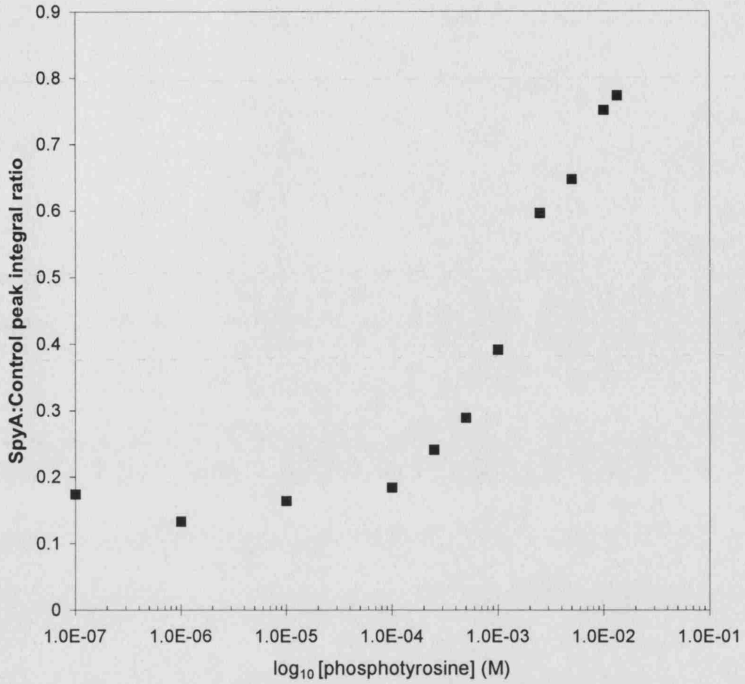


Figure 6.10 Titration of pY against the SpyA:SH2 complex leads to progressive displacement of the reporter ligand into solution as evidenced by a increase in signal intensity.

non-linearity of the relationship between $R_{2,obs}$ (and hence CPMG intensity) and the fraction of bound reporter ligand (related to [pY]), by virtue of the intermediate exchange regime that is operative for SpyA. In order to perform a robust calculation of the dissociation constant of pY (or SpyA), data points need to be fitted to the equation describing CPMG signal intensity (6.3) and thus one needs to know values for the free and bound state transverse relaxation rates and chemical shifts, and the ligand residence time constant. Due to the intermediate exchange rates in operation, it is difficult to obtain information regarding the bound state. However, without this information attempts to fit the observed data to equations describing CPMG intensities and competition equilibria will involve three unknowns: K_{i}, $R_{2,b}$, δ_f–δ_b. As shown in simulations above (Figure 6.2), alteration in the latter two variables has a significant effect on the relationship between the fraction of bound ligand and the available signal intensity. This uncertainty subjects the calculation of K_{i} to potentially large errors.

The calculation is simplified by assuming that the Spy molecule is in fast exchange with the SH2 domain. In this case, $R_{2,obs}$ is linearly dependent on the fraction of bound ligand, and the data may be fitted to a sigmoidal curve describing an apparent binding equilibrium:

$$I = I_0 + (I_{max} - I_0)[\mathrm{PL}] \quad (6.22)$$

$$\text{where } [\mathrm{PL}] = \frac{(K_\mathrm{d}^* + \mathrm{L_{tot}} + \mathrm{P_{tot}}) - \sqrt{(K_\mathrm{d}^* + \mathrm{L_{tot}} + \mathrm{P_{tot}})^2 - 4\mathrm{L_{tot}P_{tot}}}}{2\mathrm{P_{tot}}} \quad (6.23)$$

$$\text{and } K_\mathrm{d}^* = K_\mathrm{i}\left(1 + \frac{[\mathrm{L}]}{K_\mathrm{d}}\right) \quad (6.24)$$

In these equations, the dissociation constants of SpyA and pY (K_d and K_i, respectively) are related to an apparent dissociation constant, K_d^*, which is obtained by fitting the observed signal intensity data (I) in Figure 6.10. Limits for the sigmoidal curve are given by the signal intensities expected in the absence of competitor (I_0) and in the presence of an infinite competitor concentration (I_{max}). Equation 6.22 was used to fit the observed data from the titration of pY against SpyA signal intensity, yielding an apparent dissociation constant of 1.79 ± 0.21 mM (Figure 6.11). The fraction of SpyA molecule that is free will vary across the experimental samples between 0.6 and ≈1.0 (i.e. [L] = 120–200 μM)*. Using this range and the ITC-derived K_i of pY (175 ± 27 μM), the K_d for SpyA is calculated to be between 12.9 and 21.5 μM. Compared to the value obtained in ITC analysis above (177 ± 29 μM), the affinity SpyA is overestimated by approximately one order of magnitude by this simplified method.

6.3.4 Proton decoupling improves sensitivity

It was noted in 1D ^{1}H-NMR spectra that the 3'H resonance of SpyA was split into a doublet – the spin was coupled to the neighbouring CF_3 group†. Application of a ^{19}F decoupling pulse caused the 3'H resonance to coalesce into a single peak (data not shown). The reverse coupling also occurred in the ^{19}F-NMR spectrum, and therefore the CPMG pulse sequence was modified to include ^{1}H decoupling during acquisition.

The beneficial effect of ^{1}H decoupling on the ^{19}F-NMR signal intensity was explored by running a typical FAXS experiment with and without the decoupling pulse, which had a duration of 100 μs with a power attenuation of 18 dB (low power). The pulse frequency was centred at the 3'H spin resonance (~7.5 ppm) in order to achieve a high level of decoupling efficiency. The CPMG pulse sequence was used to obtain spectra of SpyA and Control in the absence and presence of v-Src

*This is calculated using Equations 6.7 and the known dissociation constant for SpyA, derived from ITC experiments.

†The alternative parallel and antiparallel orientation of the ^{19}F spins causes the 3'H resonance to be split into two peaks.

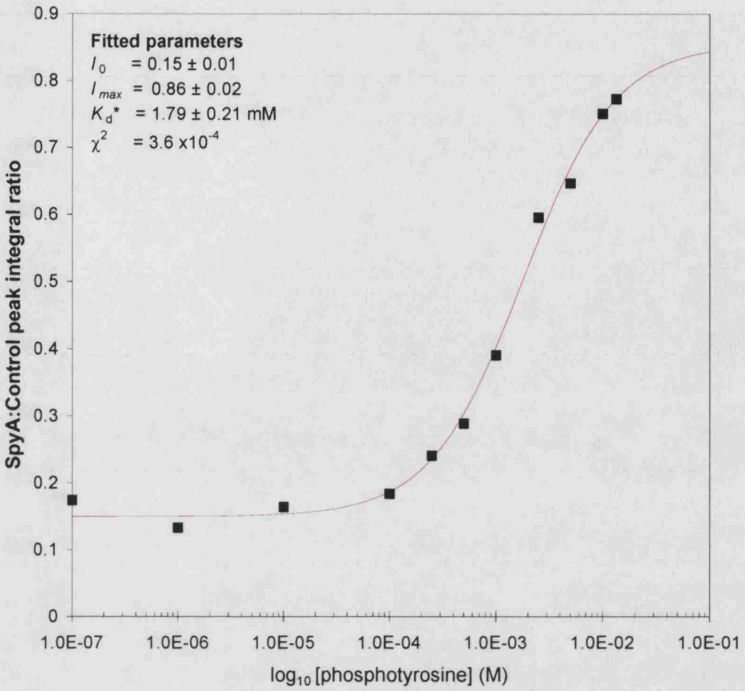


Figure 6.11 Titration of pY against the SpyA:SH2 complex leads to progressive displacement of the reporter ligand into solution as evidenced by a increase in signal intensity. These data were fitted to Equation 6.22 which are valid for fast exchange rates only. Values for the final fitting parameters are shown inset.

SH2, and subsequently in the presence of a large excess of phosphotyrosine. The results (Figure 6.12) showed that application of the ^{1}H decoupling pulse caused a 25 % mean increase in the CF_3 resonance intensity, which reflects an overall improvement in sensitivity. Proton decoupling was employed in all further CPMG experiments involving SpyA; this had a negligible effect on the total scan time.

6.3.5 Limits of detection

For efficient application of FAXS, the minimum possible concentration of reporter compounds is controlled by the sensitivity of the NMR spectrometer in detecting the ^{19}F signal. Most screening strategies employ minimal sample concentrations in order to reduce cost and avoid problems with non-specific binding and compound insolubility. However, in order to obtain robust Spy molecule peak integrals, the data must be of sufficient quality, and this is readily achieved using high molecular concentrations. Experimental conditions are therefore optimised by varying sample conditions and the number of scans until the required level of sensitivity is achieved. Dalvit *et al.*[93] demonstrated that FAXS could operate effectively at low reporter ligand[‡] concentrations (20 μM), requiring 128 scans (7 minutes) on a 600 MHz spec-

[‡]A trifluoromethyl Spy molecule was used in this report.

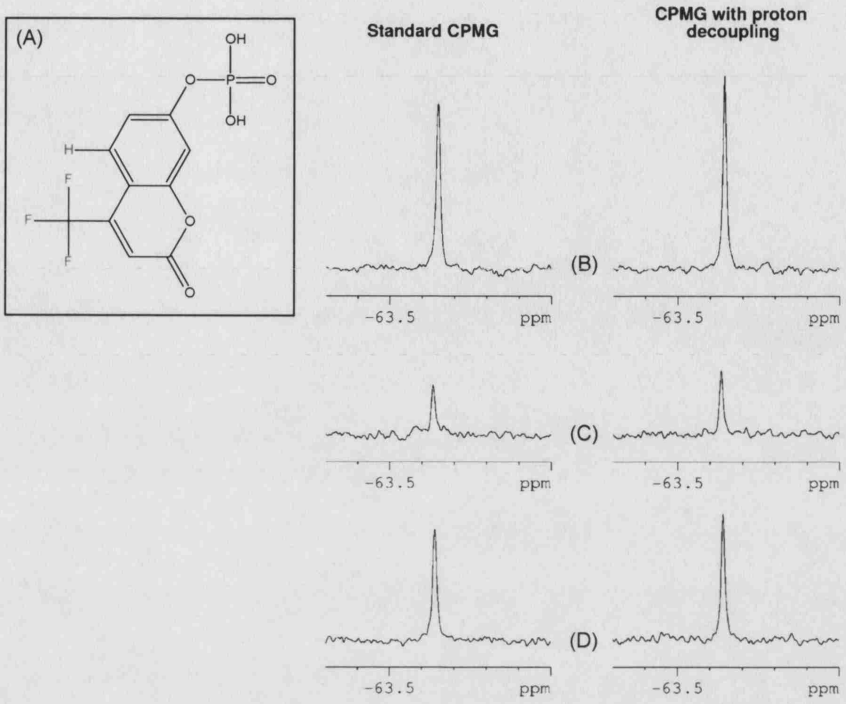


Figure 6.12 The effect of proton decoupling on competition CPMG spectra. (A) Spy molecule CF_3 group (green) is coupled to neighbouring 3'H proton (red). (B-D) CPMG spectra with (right) and without (left) proton decoupling. Each sample contains 200 μM SpyA and 50 μM Control. The Control peak is not shown, for clarity. (C) SpyA & Control in the presence of 200 μM *v*-Src SH2; (D) SpyA, Control & protein in the presence of 13.4 mM phosphotyrosine. Peak heights were used to calculate the improvements in signal intensity for SpyA.

trometer. This experiment was replicated on Celltech's 400 MHz spectrometer with the SpyA–SH2 system, using an increased number of scans to compensate for the reduced equipment sensitivity.

CPMG spectra were recorded using a sample containing 20 μM SpyA and 5 μM Control, to which *v*-Src SH2 and phosphotyrosine was added sequentially to final concentrations of 20 μM and 1.8 mM, respectively. The results (Figure 6.13) indicate that the SpyA–SH2 interaction and competition by phosphotyrosine can be observed as these low concentrations, however the signal to noise ratio is much lower compared to the equivalent experiment performed at 10-fold higher concentrations. Sensitivity is also lower than the analogous experiment performed by Dalvit *et al.* at higher field strength. The significant influence of spectral noise may interfere with the accuracy of automated phasing and integration procedures, and prevent robust measurement of competitor dissociation constants.

As described above, the SH2 domain interacts weakly with SpyA, and is required in high concentration to fully suppress the ^{19}F CPMG signal. Thus, for this system (and other fragment ligands binding to small domains), optimisation of the Spy

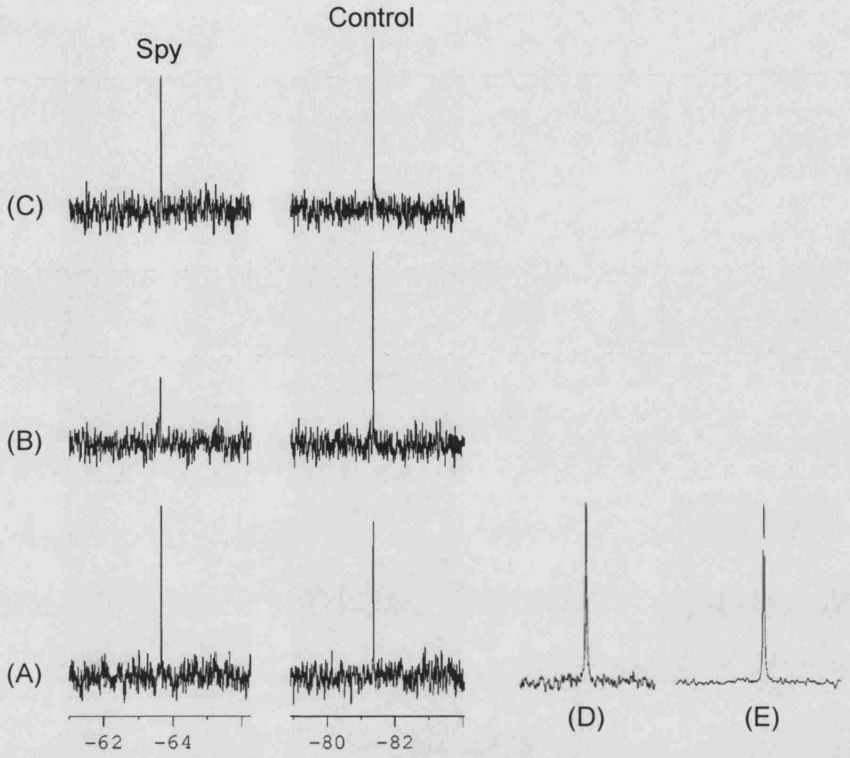


Figure 6.13 Testing the performance of FAXS at low reporter molecule concentrations and using 256 scans. Spectrum (A) represents SpyA (20 μM) and Control (5 μM) alone. Addition of protein (20 μM) to this sample yielded spectrum (B) in which signal from SpyA is selectively reduced. The presence of phosphotyrosine (1.8 mM) in the final spectrum of the series (C), indicates a strong recovery of the SpyA peak. For comparison, (D) was recorded by Dalvit *et al.*[93] using 25 μM Spy (trifluoromethyl) and 128 scans on a 600 MHz spectrometer. (E) displays the SpyA peak obtained at 10-fold higher concentration (200 μM) recording 32 scans on the 400 MHz machine at Celltech.

molecule concentration in terms of experimental sensitivity has an important effect on protein requirements and compound solubility. Since the SH2 domain is readily expressed and purified it is not unfeasible to perform FAXS using 75–100 μM reporter ligand concentration[§]. However, if test compounds are to be screened at a 5–10 fold concentration excess over reporter ligand (0.38–1.0 mM, in the scenario presented), this may lead to solubility problems and force a reduction in sample concentrations and a compensatory increase in the number of scans[¶].

[§]In this case, a screen of 1000 fragments in mixtures of 10 would require 49–65 mg of SH2 domain. This calculation is based on 100 samples, each 500 μL, with [SH2] = 75–100 μM.

[¶]The time required to perform 100 experiments employing 128, 192, or 256 scans is 14.2, 21.3, or 28.3 hours, respectively. This does not include the time needed for sample loading, equilibration, and adjustment of shim coils, which is approximately 2-3 minutes per sample.

6.4 Screening of phosphotyrosine mimetics using FAXS

The phosphotyrosine mimetics fragments identified through computational methods (Chapter 5) were screened against *v*-Src SH2 using the optimised FAXS experiment. Although fourteen compounds were initially identified, four could not be purchased, and three failed to arrive in time for this screening analysis. The remaining seven fragments were screened, and compared to results for three known binders (pY, phenylphosphate, and naphthylphosphate), and the two alternative reporter ligands (SpyB and SpyC). Fragments were tested individually at a concentration of 1.5 mM, in the presence of equimolar SpyA and SH2 domain (200 μM) and Control (50 μM). At these concentrations, 32 scans provided spectra of sufficient quality, portions of which are displayed in Figure 6.14.

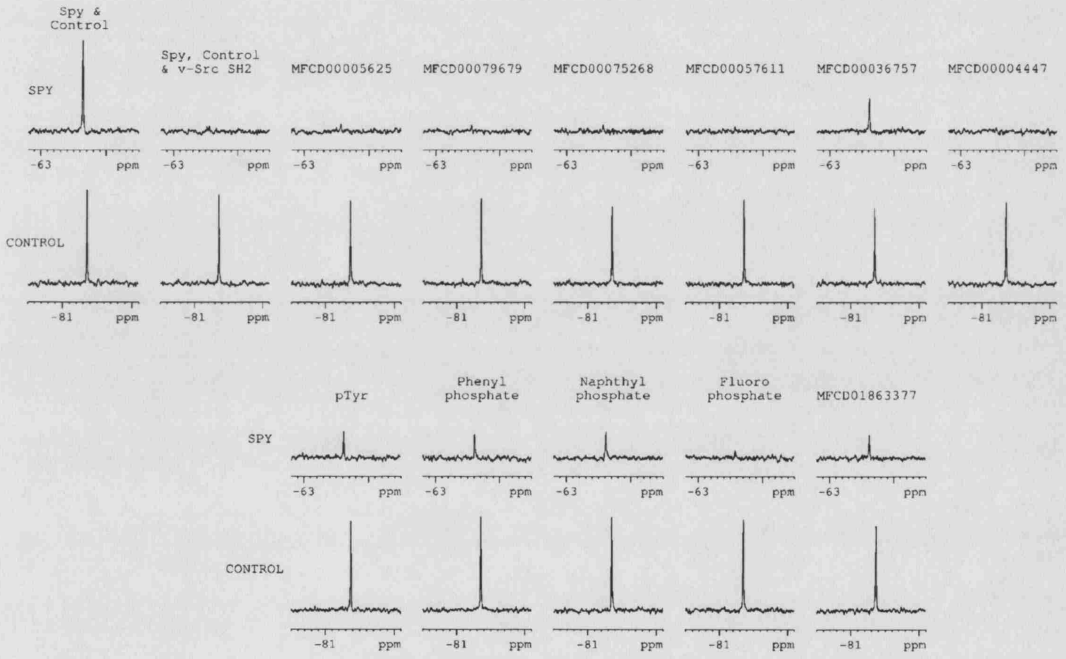


Figure 6.14 Results of FAXS screening of known and potential phosphotyrosine mimetic fragments. The SpyA and Control resonances were isolated from each spectrum and stacked vertically. Thus, the first and second rows of spectra indicate the effect of each ligand on SpyA and Control, respectively. This pattern is repeated for the third and fourth rows. Starting from the top left, the first two experiments were performed in the absence and presence of *v*-Src SH2 domain. Screening experiments were then performed as described in the main body of text. Fragments structures can be identified by their MFCD number on page 216.

Screening of compounds indicated that two fragments novel to *v*-Src SH2 (MFCD numbers 36757 & 1863377 (SpyC)) competed with SpyA, which strongly suggests that they interact with the domain. Both of these fragments feature an aromatic–phosphate moiety and are therefore closely related to phosphotyrosine. The remain-

ing five compounds failed to compete with the spy molecule for the SH2 binding site, however it remains possible that they have very weak affinity for *v*-Src SH2. The known binders – phenylphosphate and naphthylphosphate – indicated competitive abilities analogous to phosphotyrosine, however the weak-binding SpyB (FPO_3) fragment showed essentially zero competitive ability. From analysis of SpyA peak integrals, it appears that MFCD00036757 (hereafter referred to as 36757) is a slightly more efficient competitor than the known ligands. Binding of this novel fragment ligand was explored further by NMR and ITC (see below).

6.4.1 Testing a mixture of compounds

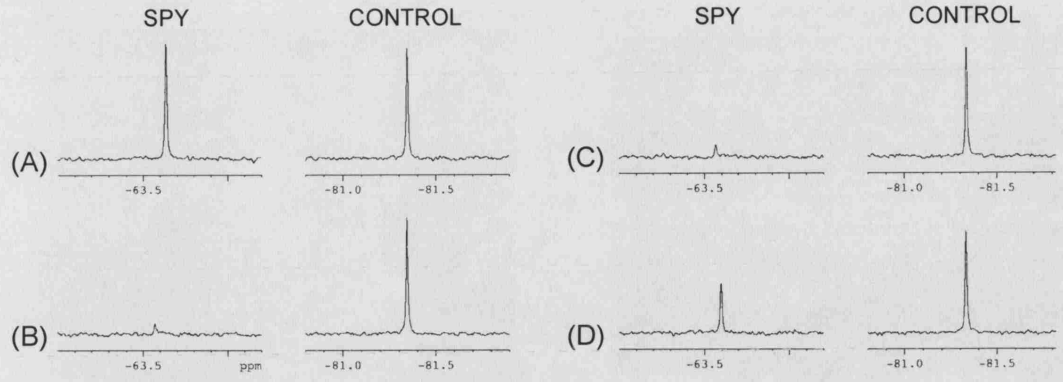
In order to maximise the throughput of screening, compounds are often tested in mixtures. However, methods that do not monitor observable parameters for each screening compound (e.g. biological assays, competition NMR approaches) require follow-up mixture deconvolution experiments to identify the active ligand. The FAXS experiment was performed with the SH2 domain and a small mixture of non-binding compounds (1 mM each) in the absence and presence of a known ligand (36757, 1 mM). The results indicated that the mixture of non-binding fragments had no effect on the NMR signal of SpyA, which suggests that they did not interfere with the SpyA–SH2 interaction (Figure 6.15). Addition of the known binder to the mixture resulted in competitive displacement of SpyA to levels previously observed for 36757. Thus FAXS can be used to screen of mixtures of compounds, however these should be carefully designed such that mixture deconvolution is efficient where competition is observed[‖].

6.4.2 Use of SpyB (fluorophosphate) as reporter ligand

Since phosphate ions interact weakly with the phosphotyrosine pocket of *v*-Src SH2 (see Section 4.1.1), fluorophosphate (FPO_3) was selected as an alternative spy molecule for FAXS experiments. Competition ITC analysis provided a thermodynamic description of the interaction between FPO_3 and the SH2 domain (see Section 6.2.2 above) and indicated a dissociation constant of 1.63 ± 0.21 mM, which is approximately an order of magnitude weaker than SpyA and phosphotyrosine. Indeed, as Figure 6.14 shows, a 7.5 molar excess of FPO_3 was not able to cause significant displacement of SpyA. Thus, it is possible that other fragment compounds which failed to elicit a response in screening against SpyA may compete more effectively against SpyB.

From a sensitivity point of view, fluorophosphate is a sub-optimal choice due

[‖]The optimal size of NMR screening mixtures was discussed recently by Mercier & Powers[334].



Non-binders: MFCD numbers 4447, 5625, 57611, 75268 and 79679, pyruvate, clozapene, glutamine, methionine and sucrose.

Known binder: MFCD00036757

Figure 6.15 Screening a mixture of compounds using FAXS. Four samples were prepared and subjected to the CPMG experiment. Each contained SpyA (200 μM) and Control (50 μM). Spectrum (A) shows the isolated resonances for free SpyA and Control molecules. Addition of 200 μM *v*-Src SH2 lead to a significant and selective reduction in the SpyA peak integral (B). The third sample (C) also contained ten non-binding fragment compounds (named above) at a concentration of 1 mM each. Finally, a known binder was included in the mixture, and resulted in 60 % recovery of the SpyA peak (D).

to the existence of a heteronuclear coupling between the fluorine and phosphorus, which leads to splitting of the ^{19}F signal into a doublet (Figure 6.16), and a reduction of the signal to noise ratio by 50 %. Whilst decoupling pulses can be applied to resolve this situation, the most common cases involve either decoupling protons from heteronuclei, or *vice versa*. This rare case in which ^{31}P needs to be decoupled from ^{19}F was not achievable at Celltech due to equipment limitations. Thus the lower signal of this monofluorinated spy molecule was reduced further by splitting into a doublet.

6.4.2.1 Detection of SpyB binding with FAXS

The 90°_x and 180°_y pulse widths relevant to SpyB were optimised as described above for SpyA (Section 6.3.1). CPMG spectra of SpyB and Control were recorded in the absence and presence of *v*-Src SH2 using the same relaxation delay period employed above (i.e. τ = 20 ms, n = 8). Due to the lower sensitivity of SpyB, 256 scans were required to obtain data of sufficient quality, which was treated with 3 Hz line broadening prior to Fourier transformation. Comparison of these spectra (Figure 6.17 A & B) indicated a significant decrease in SpyB peak intensities for the sample containing SH2 domain, whilst the Control molecule peak integral remained

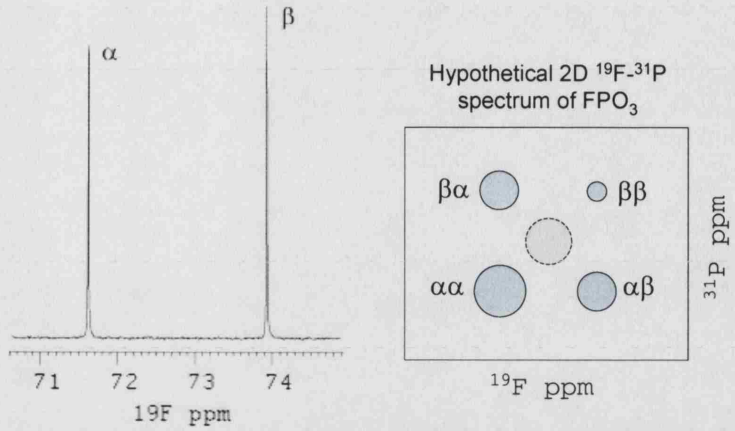


Figure 6.16 1D ^{19}F-NMR spectrum of FPO_3 (SpyB). The ^{19}F resonance is split into a doublet due to the alternative magnetic fields imposed on this nucleus by the α and β spin orientations of ^{31}P. Moreover the relative linewidths of the two peaks are different because the relaxation rate of each ^{19}F nucleus is modulated by the relative orientation of the attached ^{31}P spin. The local field at the ^{19}F nucleus contains a contribution from its own CSA and the ^{31}P magnetic moment. Where both spins are orientated antiparallel (β) to the external field, these contributions interfere with each other and cause a reduction in the local field strength, relaxation rate, and hence linewidth. Conversely, if both spins are in the α state, the CSA and dipolar contributions to relaxation are additive, leading to broadening of the resonance. For $\alpha\beta$ and $\beta\alpha$ combinations, intermediate linewidths are observed. In the 1D ^{19}F-NMR spectrum recorded for SpyB, the degenerate $\beta\beta$ and $\alpha\beta$ combinations generate a narrower linewidth than the $\alpha\alpha$ and $\beta\alpha$ combinations. A hypothetical ^{19}F–^{31}P 2D spectrum is shown in which the FPO_3 signal is split into a quartet, and individual peaks are labelled with the spin orientations of the ^{19}F and ^{31}P nuclei. The size of each circle is representative of the difference in linewidth associated with spin orientations. TROSY-based[96] pulse sequences exploit this phenomenon by selectively observing the $\beta\beta$ peak, which displays the narrowest linewidth. If the ^{31}P nucleus was decoupled by an appropriate pulse, the quartet would collapse into a central peak indicated by the grey circle, with linewidth approximately equal to that of the $\alpha\beta$ and $\beta\alpha$.

unaffected. This result confirms that the interaction between SpyB and *v*-Src SH2 can be detected using CPMG spectra.

The fact that eight CPMG cycles are sufficient to suppress the NMR signals of both SpyA and SpyB indicates that in the presence of receptor, $R_{2,obs}$ is similar for each ligand. Given the known dissociation constants (and hence expected bound state time constants) for each molecule, it appears that SH2–mediated enhancement of R_2 is more efficient for SpyB than SpyA. It is likely that the difference in free and bound state chemical shifts are significantly greater for SpyB than for SpyA, which despite the increased exchange rate, leads to a larger net contribution to $R_{2,obs}$ from chemical exchange (see Equation 6.5 on page 222). This probably arises because in the SpyB–SH2 complex, the single ^{19}F nucleus is within the core of the binding site and would experience a significant change in electromagnetic environment. Con-

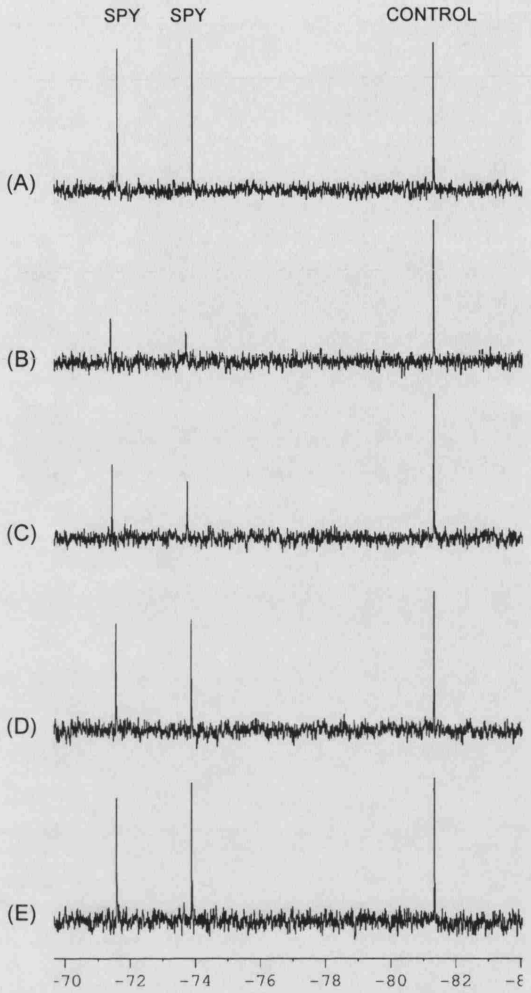


Figure 6.17 1D ^{19}F CPMG spectra showing the interaction between SpyB and *v*-Src SH2 and the effect of adding a competitive ligand. (A) SpyB (200 μM) and Control (10 μM). (B) Equimolar SpyB and SH2 domain (200 μM) and Control (10 μM). Spectra (C)-(E) indicate the effect of competitor (pY) concentration (0.1, 1.0, and 2.0 mM, respectively) on the SpyB signal.

versely the CF_3 group of SpyA is predicted to occupy a surface-exposed position that is separated from the phosphate group by virtue of the fragment chemical structure (see page 228). These differences indicate how the nature of the reporter ligand and its interaction with receptor, can have a strong influence over the experimental setup.

Interestingly, the relative sizes of the fluorophosphate doublet peaks also changed on binding to the SH2 domain. In the absence of protein, the peak at -73.9 ppm (β) was of greater intensity than its counterpart. On addition of protein, the situation was reversed, the reason for which is unknown. Addition of a relatively high concentration of phosphotyrosine caused the β resonance of SpyB to regain its superior intensity. Interaction with the SH2 domain could have a differential effect on the

relaxation rates of the ^{19}F nucleus, depending on whether it is bonded to an α or β ^{31}P spin.

On comparing the SpyB peak integrals from spectra (A) and (E) in Figure 6.17, 2 mM pY caused an 85 % recovery in signal intensity. In similar conditions, 2.5 mM pY was only able to effect a 60 % recovery in SpyA intensity (Figure 6.10). This difference confirms that *v*-Src SH2 has a higher affinity for the trifluoromethyl reporter ligand than SpyB, which supports results from ITC and FAXS screening (see Figures 6.7 and 6.14).

6.4.2.2 Re-screening of fragments using SpyB

The fragment compounds identified from computational modelling were re-screened against *v*-Src SH2 using SpyB as the reporter ligand. The concentrations of SpyB and SH2 domain were increased to 300 μM to improve signal strength, and each compound was tested individually at a concentration of 2 mM. The results of this screening were very similar those obtained when SpyA was used as the reporter (see Figure 6.18). The only novel ligand able to compete with SpyB was 36757; the other fragments displayed essentially no effect confirming that they do not interact with the SH2 domain. Interestingly, differences between the competitive abilities of the three known ligands were more apparent when screening with SpyB. The pattern of results suggests that their K_d values may be ranked as follows: phenylphosphate > naphthylphosphate > pY. Hajduk *et al.* screened these three fragments against the SH2 domain of Lck, and obtained K_d values of 0.3, 0.2, and 0.4 mM, respectively[173]. Thus, the hierarchy of fragment binding affinities appears to be slightly different for *v*-Src SH2. Given the additional size and hydrogen bonding capabilities of pY, it may be expected that this fragment would display the highest affinity of the three.

The successful use of SpyB in the identification of fragment ligands for the SH2 domain indicates that even a weak-affinity reporter ligand with sub-optimal spectral features can still be used in the FAXS experimental approach. The sensitivity of this approach would benefit from decoupling the ^{31}P nucleus from the ^{19}F signal.

6.4.3 Use of di-fluorinated spy molecules

The fragment compound MFCD01863377 (SpyC) was shown to compete with SpyA for SH2 domain binding sites (Figure 6.14, page 239). Since this ligand is fluorinated, it was put forward as a potential spy molecule for use in FAXS experiments. However, analysis of standard (i.e. pulse–acquire) 1D ^{19}F spectra of SpyC revealed complex lines shapes which are the result of mutual couplings between the fluorine nuclei, and with neighbouring protons (Figure 6.19). Proton decoupling reduced the

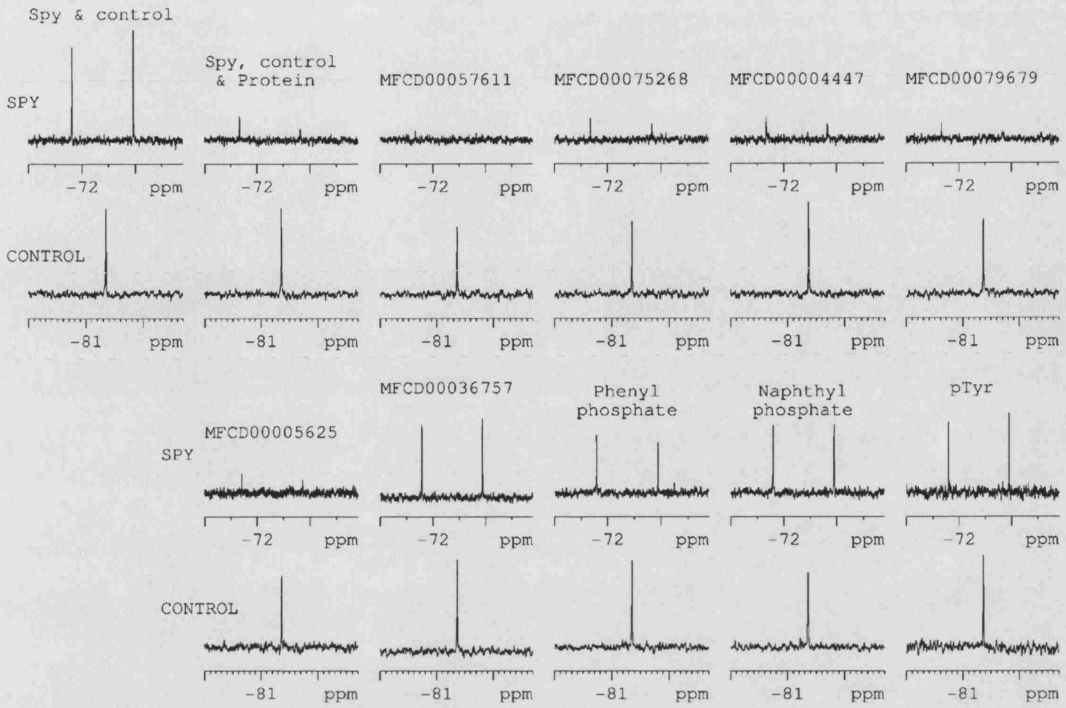


Figure 6.18 Results of FAXS experiments that utilised SpyB as the reporter ligand. Each spectrum was recorded using 128 scans (9 minutes acquisition time). The format of this Figure is identical to that described in Figure 6.14. Starting from the top left, the first two spectra are of SpyB and Control in the absence and presence of receptor. The remaining spectra indicate the effect of adding the named compounds to this ternary mixture. Only 36757 and the known binders are able to compete with SpyB.

overall spectral complexity, however due to limitations in probe design, simultaneous application of ^{1}H and ^{19}F decoupling pulses was not possible. These complications prevented use of SpyC as a reporter ligand in FAXS. Moreover, given the structure of SpyC, it is questionable whether the two fluorine nuclei would undergo large chemical shift changes on binding to SH2 domain since they are attached to a freely rotatable benzene ring and located several Å from the phosphate group. This also inhibits successful use of SpyC in the FAXS experimental system.

6.5 FAXS with a Celltech pharmaceutical target

FAXS was also demonstrated using a target protein of interest to Celltech, which will be referred to as Protein-X for intellectual property reasons. Protein-X is a ~50 kDa polypeptide which exists in solution as a homotetramer. Previous ELISA-based *in vitro* screening experiments at Celltech identified a CF_3-containing fragment with an IC_{50} value of 10 μM. This compound, termed CT_{spy}, was used as the reporter in FAXS experiments, whilst a known high-affinity ligand (IC_{50} in the nM range) was

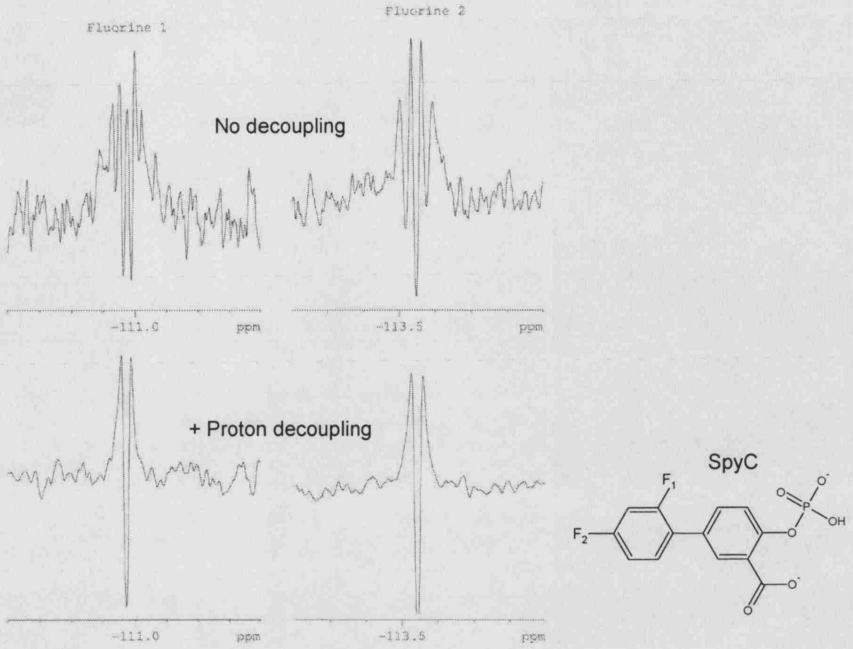


Figure 6.19 1D spectra of SpyC indicated multiple couplings to other nearby nuclei. This was partially resolved by application of a proton decoupling pulse, however the ^{19}F homonuclear coupling could not be suppressed.

employed as the competitor, CT_{com}. Such compounds are known to interact with Protein-X in a deep cavity within the receptor surface.

The CPMG experiment was used to demonstrate an interaction between Protein-X and reporter ligand, using 50 μM CT_{spy} and 10 μM Control (hexafluoroacetone). A protein concentration of 7.5 μM was sufficient to completely suppress signals from CT_{spy} (Figure 6.20), which compared with the SH2–fragment system is relatively low. The ~7-fold reduction in protein requirements indicates that Protein-X is more efficient at inducing relaxation of reporter ligand nuclei than *v*-Src SH2. This can be rationalised in several ways, with reference to Equations 6.3–6.5 describing FAXS on page 222. Firstly, the effective molecular weight of ~200 kDa makes Protein-X about 20-fold larger than the SH2 domain. The rotational correlation time τ_c is therefore longer, which leads to a significant increase in the bound state transverse relaxation rate, $R_{2,b}$. Whilst the K_d and IC_{50} data for SpyA and CT_{spy} are not directly comparable, a higher affinity of interaction with Protein-X would tend to decrease the off rate and cause the reporter ligand to associate with the receptor for longer periods, which also increases $R_{2,obs}$. Thirdly, it is possible that there is a greater chemical shift difference between free and bound CT_{spy} molecules than for the analogous SpyA–SH2 interaction, which would contribute significantly to $R_{2,obs}$. This is likely given the fact that CT_{spy} binds within a deeper cavity than SH2; the CF_3 group may also be less exposed to solvent. The CT_{spy} reporter ligand

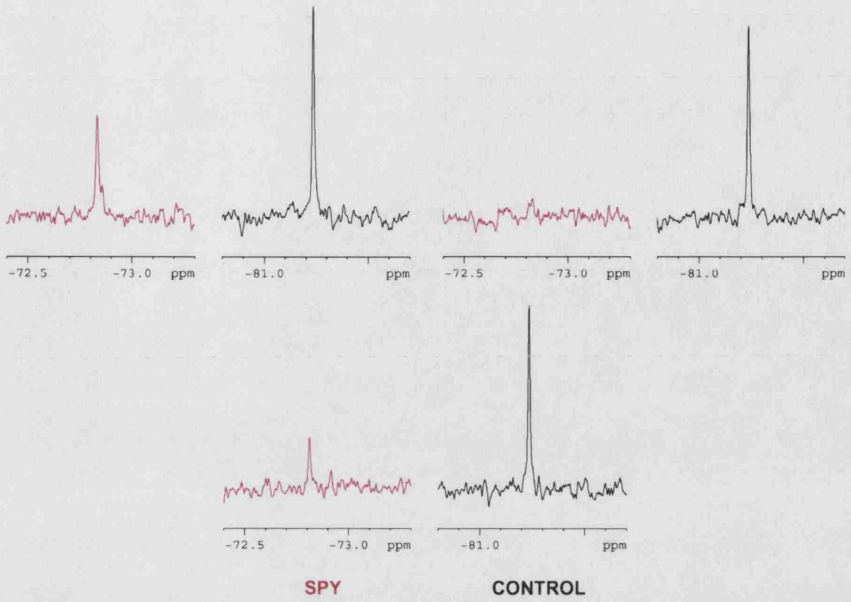


Figure 6.20 Demonstrating the FAXS experimental approach with a target protein of current pharmaceutical interest. The CF_3 resonances of CT_{spy} and Control are displayed in red and black colours, respectively. CPMG spectra were recorded for three samples, each containing 50 μM CT_{spy} and 10 μM Control. The first and second spectra were obtained for reporter compounds in the absence (*top left*) and presence (*top right*) of 7.5 μM Protein-X, respectively. The third sample (*bottom*) also contained Protein-X and 500 μM CT_{com}, the competitive ligand. CPMG pulse widths were optimised prior to acquisition and the relaxation delay parameters were equivalent to those used for the SH2 system (i.e. $n = 8$, $\tau = 20$ ms).

was partially displaced from Protein-X binding sites by addition of a 10-fold molar excess of CT_{com} (Figure 6.20).

6.6 Confirmation of screening results

Computer-based research led to the selection of several fragments that were chemically similar to known SH2 domain ligands and were predicted to bind to the *v*-Src SH2 domain (Chapter 5). These were screened *in vitro* using the FAXS experiment, which was optimised for weak-binding compounds, on a 400 MHz spectrometer. Of the eight novel fragments analysed, three displayed affinity for *v*-Src SH2: SpyA (800300); SpyC (1863377); and 36757. Moreover, fluorophosphate (SpyB) indicated weak affinity, and is also novel to the drug design literature relevant to the SH2 domain.

Screening results were validated by monitoring the effect of fragment compounds on SH2 domain backbone amide chemical shifts using the $^{1}H^{15}N$-HSQC pulse sequence. A 2D spectrum was first obtained for apo ^{15}N-labelled *v*-Src SH2 (0.5 mM).

Spectral parameters and processing details were identical to those employed previously for chemical perturbation studies (see Sections 2.3.2 & 4.1.1). The apparent non-binding compounds (4447, 5625, 57611, 75268, 79679, and Control) were then added to the sample to final concentrations of 1.8 mM each, however no changes in SH2 domain chemical shifts were observed (data not shown). This provides further confirmation that these fragments do not interact with the *v*-Src SH2 domain, and were thus incorrectly identified by computational methods.

The binding behaviour of two fragment ligands, SpyA and 36757, were also analysed using the $^{1}H^{15}N$-HSQC experiment. Due to structural similarities with phosphotyrosine, it was expected that these two fragments would cause similar chemical shift changes to those described for the native ligand in (Section 4.1.2), and that any differences might aid in the interpretation of predicted binding orientations and observed thermodynamic data.

6.6.1 SpyA titration

$^{1}H^{15}N$-HSQC of *v*-Src SH2 recorded at increasing concentration of SpyA are shown overlaid in Figure 6.21. The broadening of amide resonances during the middle stages of the titration indicated that SpyA is in intermediate exchange with the SH2 domain. Absolute differences in amide chemical shifts between the apo and SpyA-bound forms of *v*-Src SH2 were calculated in the same manner as for previous titrations**.

The pattern of chemical shift changes induced by SpyA binding matches closely that associated with pY (compare Figure 6.21 above with Figure 4.3, page 154). This strongly implies that both fragments occupy the same binding site and engage with the SH2 domain surface in similar orientations. If SpyA assumed an alternative binding mode, the ring current effects associated with this fragment would have had led to significant differences in chemical shifts compared to the pY-bound state. The magnitude of combined shift changes were used to colour the peptide-bound structure of *v*-Src SH2, which was used in FLEXX docking calculations (Figure 6.22). The location and magnitude of shift changes are generally consistent with the predicted binding mode of SpyA, which provides further weight to the validity of docking results. These observations support previous thermodynamic analysis in which SpyA and pY appear to interact with the SH2 domain with similar energetics and affinities (Section 6.2.2). Despite the overall similarity between the results, the presence of the umbelliferyl moiety (bicyclic structure) of SpyA is evident from the altered behaviour of nearby backbone amides. For example, K60 displays a large and negative ^{1}H shift change in the SpyA complex whereas pY induces a similarly large positive

**Combined shift change = ($|\Delta^{15}N| / 6.4$) + $|\Delta^{1}H|$

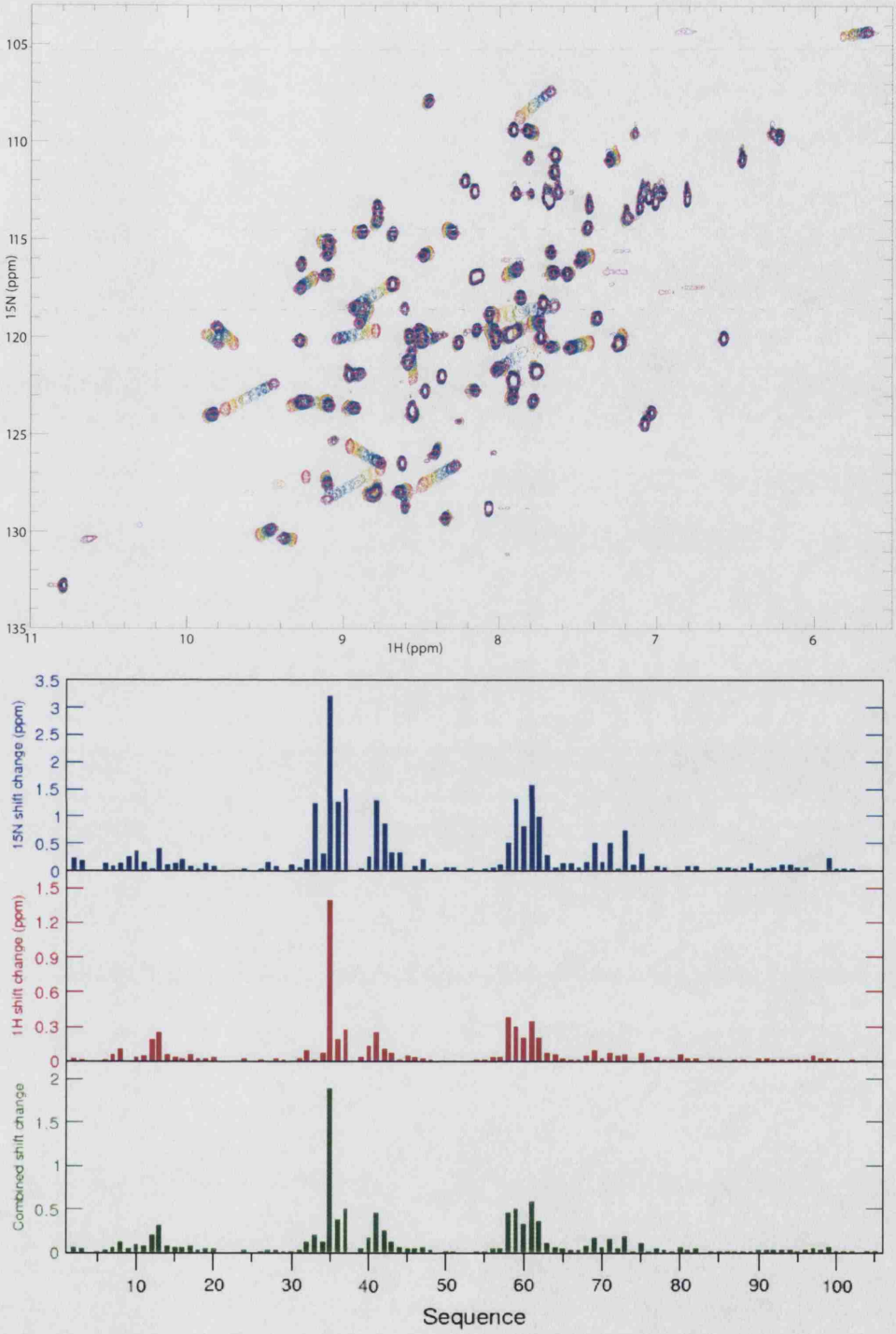


Figure 6.21 Results of SpyA titration. *Top* | Overlaid $^{1}H^{15}N$-HSQC spectra describing the progressive titration of SpyA into apo *v*-Src SH2 domain. Contours are coloured according to the visible light spectrum, from red (apo) to indigo (saturated). *Bottom* | Chemical shift differences between free and bound state are depicted as raw ^{15}N (blue) and ^{1}H (red) data, as well as in the combined format (green).

change. Additionally, the change in the ^{15}N shift of Y59 is considerably smaller on binding SpyA compared to pY.

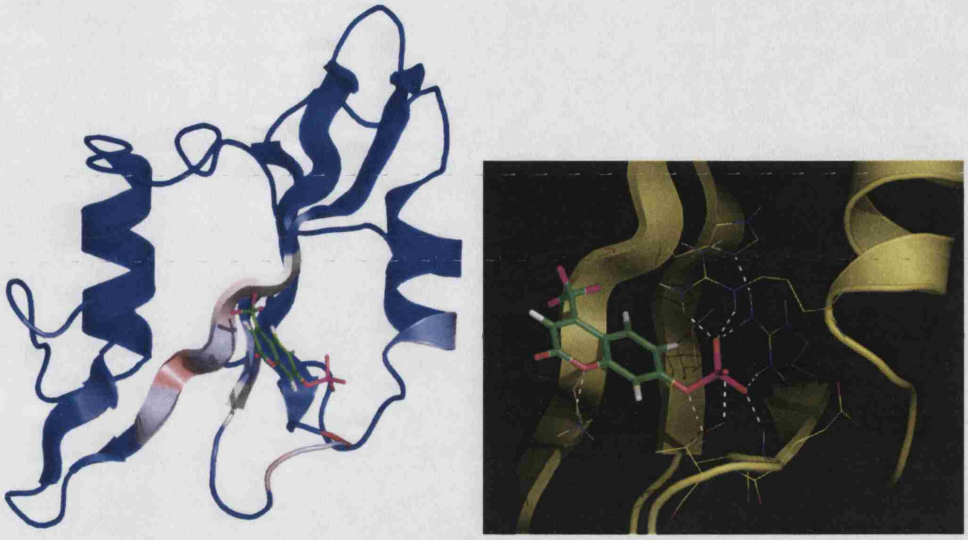


Figure 6.22 Analysis of SpyA-induced backbone amide chemical shift changes using the predicted structure of the complex as a model. *Left* | The *v*-Src SH2 structure is coloured according to the extent of local shift change. *Right* | This image shows the hydrogen bonding interactions (white dashes) that are predicted to occur at the SpyA/SH2 interface.

6.6.2 36757 titration

The ^{1}H^{15}N-HSQC titration procedure was also used to analyse the effects of 36757 on *v*-Src SH2 chemical shifts. Unfortunately the protein sample displayed evidence of slight degradation and the resultant spectra (see Figure 6.23) were not as clean as for the other titrations. The characteristic shift in the E35 amide did occur during the titration but this resonance did not re-appear in the spectrum suggesting that 36757 may induce some form of slow-intermediate exchange process in that part of the BC loop and may be an artefact of the poorer sample quality. Alternatively, the binding ability of the domain may be somewhat abrogated by degradation and the titration may not have reached saturation. Despite these issues, which can be overcome by use of fresh samples, the interaction of 36757 with *v*-Src SH2 clearly occurs at the pY binding site as it induces shift changes that are localised to this region and follow a very similar pattern to pY and SpyA fragment ligands.

6.6.3 ITC analysis of 36757

The competition ITC technique described in Section 6.2.2 was used to measure the ΔH_b° and K_d of the screening hit, 36757, and enable calculation of ΔS_b° and

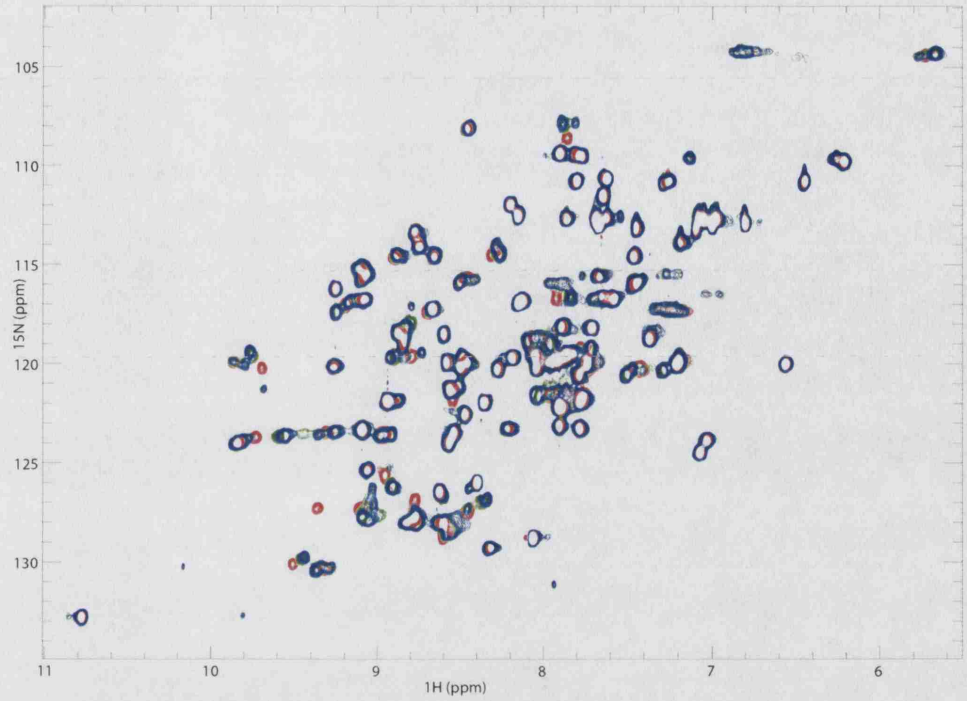
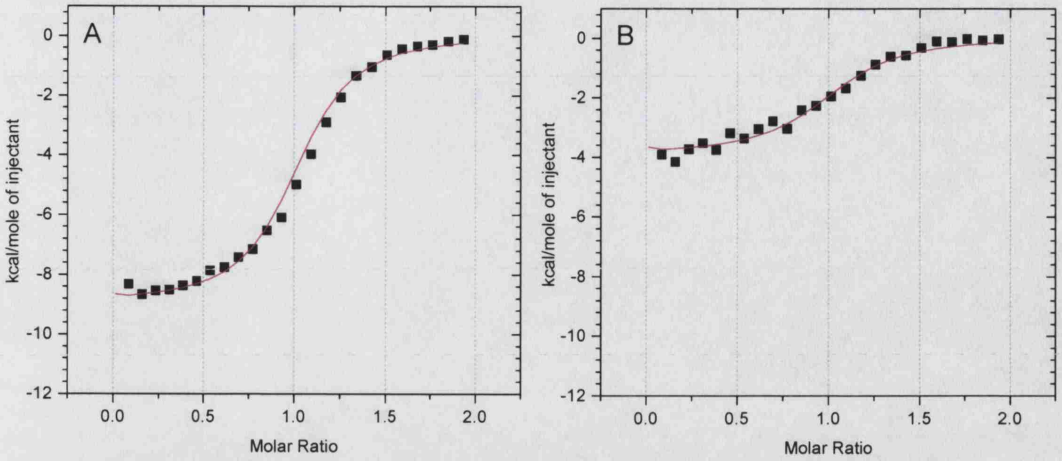


Figure 6.23 Overlaid $^{1}H^{15}N$-HSQC spectra describing the progressive titration of 36757 into a sample containing *v*-Src SH2. The contour colours are red: apo domain (0.5 mM); green: +0.5 mM 36757; and blue: +1.5 mM 36757.

ΔG_b°. The experimental conditions for this analysis were identical to those employed previously (i.e. 35 μM *v*-Src SH2 and 2 mM 36757 in the calorimeter cell, 0.24 mM peptide in the syringe). The apparent binding isotherm of PQpYEEIPI peptide in the presence of 36757 is shown in Figure 6.24 along with the analogous result for phosphotyrosine. Full thermodynamic data describing the interaction between 36757 and *v*-Src SH2 are also provided along with data for pY and SpyA.

The thermodynamic profiles of 36757, pY, and SpyA are significantly different. The interaction of the SH2 domain with 36757 is dominated by a large, negative change in ΔH_b° and small positive change in ΔS_b° whilst the opposite pattern is observed for pY and SpyA fragment ligands. The more negative (i.e. more favourable) ΔH_b° change associated with 36757 probably represents an increase in the strength and/or number of intermolecular interactions made with the SH2 domain compared to the SH2–pY complex. The less positive (i.e. less favourable) change in $T\Delta S$ observed with 36757 compared to pY and SpyA suggests an additional restriction of molecular freedom for the protein and/or solvent. It is possible that solvent molecules are involved in formation of the SH2–36757 complex, however given the small size of the fragment and the lack of empirical structural information, this seems unlikely and cannot be substantiated. Interestingly, the favourable gain in enthalpy observed for 36757 is counteracted by the unfavourable loss of entropy such



Fragment	K_d (μM)	ΔH_b° (KJ.mol^{-1})	TΔS (KJ.mol^{-1})	ΔG_b° (KJ.mol^{-1})
36757	**104 ± 28**	**-19.04 ± 0.56**	**3.68**	**-22.72**
pY	175 ± 27	-3.30 ± 0.49	18.13	-21.43
SpyA	177 ± 29	-5.53 ± 0.53	15.87	-21.40

Figure 6.24 Thermodynamic analysis of 36757 binding. *Top* | Apparent binding isotherms for the PQpYEEIPI ligand in the presence of pY (*left*) and 36757 (*right*). Full thermodynamic data is provided for 36757 and compared to data obtained previously for pY and SpyA.

that the overall free energy describing the interaction of the SH2 domain with each fragment is very similar.

The binding conformations of 36757 and pY suggested by FLEXX (Figure 6.25) indicates that the phosphate group of both ligands makes an identical set of hydrogen bonding interactions with the core of the pY pocket. The aromatic rings of both fragments are also predicted to occupy similar positions, and it seems likely that π-interactions with R12 and/or K60 would occur in both complexes. Differences in binding thermodynamics for these two fragments may therefore be attributed to the three aromatic substitutions that distinguish 36757 from pY, that is, the indole amide, bromine, and chlorine. In the predicted structure of the SH2–36757 complex, the two halogen functionalities are orientated away from the surface, and are exposed to solvent. The indole amide however is close to the domain surface and could form a hydrogen bond with the nearby carbonyl oxygen of His βD4 (Figure 6.25). In the 'static' structure of the complex, this appears to be unlikely due to the sub-optimal distances and angles between the amide and the carbonyl. However, in reality, a hydrogen bond could be formed if one or both groups were moved slightly. An 'induced fit' hypothesis may therefore explain the observed thermodynamic data: The additional hydrogen bond would be expected to have a favourable effect on ΔH_b° but reduce TΔS by virtue of the additional strain associated with distortion of the backbone at His βD4.

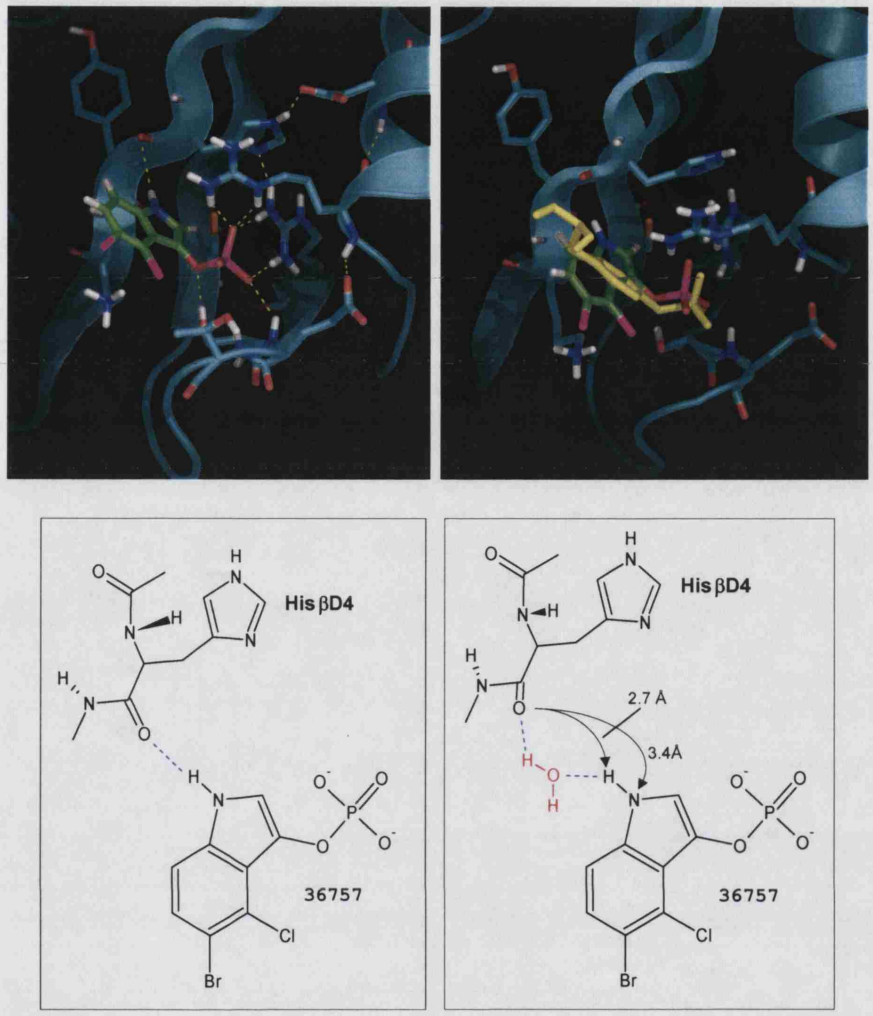


Figure 6.25 Predicted binding conformations of 36757 and pY, as determined by FLEXX. *Top left* | Potential hydrogen bonding interactions between 36757 and the SH2 domain are indicated by yellow dashes. Note the additional hydrogen bond suggested between the indole amide of 36757 and carbonyl oxygen of His βD4. *Top right* | The docking orientations of 36757 (atom colours) and pY (yellow sticks) are overlaid for comparison. *Bottom left* | A conceptual model of the hydrogen bond that may exist between the indole amide of 36757 and His βD4. *Bottom right* | An alternative model that invokes a bridging water molecule.

If the energetic penalty associated with backbone deformation is too high, it is possible that a water molecule could provide bridging hydrogen bonding with the indole amide of 36757 (illustrated in Figure 6.25). This explanation is also compatible with the thermodynamic data as the trapping of a solvent molecule in the complex would have an unfavourable effect on $T\Delta S$. Indeed, the SH2–PQpYEEIPI complex is known to include up to four bridging water molecules (Figure 1.16), and is also associated with a small $T\Delta S$ term (5.1 KJ.mol^{-1}, see Figure 6.7).

Distinguishing between these two hypotheses requires additional data. A high

resolution structure of the complex would reveal the actual binding conformation of 36757 and validate the accuracy of the FLEXX prediction. Mass spectrometry could be used to explore the molecular weight of the complex. Removal of the indole functionality of 36757 would also provide conclusive data. If a water molecule is being employed at the interface then substitution of the indole amide with a moiety capable of forming optimal hydrogen bonding interactions directly with HisβD4 could have a favourable effect on TΔS and therefore ΔG°_{b}. Further use of 36757 in the development of an SH2 inhibitor is discussed in Chapter 7.

6.7 Extension of FAXS to phosphorus nuclei

During the course of this research we considered whether the FAXS approach could be replicated using ^{31}P-labelled reporter compounds (i.e. 'PAXS'). Although the natural abundance of ^{31}P is 100 %, this nucleus is inherently less sensitive than ^{19}F and ^{1}H (6.6 % relative to ^{1}H). However, the chemical shift range of ^{31}P is slightly larger than ^{19}F (430 and 400 ppm, respectively), and thus the contribution of chemical exchange to $R_{2,obs}$ should remain considerable. The low sensitivity of ^{31}P may prevent general use of a PAXS screening approach. However nature of the target protein and its biological interactions may provide the necessary justification for employing ^{31}P nuclei. A significant number and diversity of proteins are known to interact with phosphate-based ligands, such as metabolic intermediates, enzyme cofactors, nucleotides, nucleic acids, lipids, modified protein motifs, or carbohydrates. The natural ligands for these receptors could be used as reporter ligands in PAXS with little or no modification. For example, phosphotyrosine could act as the Spy molecule when screening against SH2, PTB, or PTP targets. This not only circumvents the requirement to fluorinate ligands for FAXS, but it also means that screening can be carried out in competition with the natural compound. Alternatively a ^{31}P-based compound could be selected from an existing database. A cursory examination of the ChemACX database* indicated 7235 compounds (2.9 %) containing at least one phosphorus atom (Dr Will Pitt, personal communication).

In many cases, the phosphate moiety is crucial to the interaction of organophosphate ligands with receptors by providing a significant proportion of the binding energy through charge-assisted hydrogen bonding. Ligand binding is therefore likely to cause relatively large perturbations in chemical shift. Indeed, the ^{31}P shift of pY varied by ~2.0 ppm over the pH range 2.9–8.6†. The high sensitivity of ^{31}P to the surrounding electromagnetic environment renders it particularly useful for R_2-edited

*This version of the Available Chemical Database is compiled by *CambridgeSoft*.

†These data were used to calculate the phosphate pK_a of PQpYEEIPI in the free and bound states as described in Section 4.3.7

screening. These possibilities prompted further investigation, using the SH2 domain as a test system.

6.7.1 Effect of binding on phosphorus linewidth

Since the linewidth of a given resonance is proportional to πR_2, the appearance of small molecule NMR spectra has been used for many years as evidence for binding to macromolecules[68]. Comparison of 1D ^{31}P spectra of phosphotyrosine (2 mM) in the absence and presence of *v*-Src SH2 (0.7 mM) indicated a significant increase in linewidth (data not shown). This can be explained by rapid transverse relaxation through reversible interactions with the domain. Interestingly, this obvious effect was not noted for interaction of SpyA with *v*-Src SH2 in equivalent ^{19}F spectra, suggesting that pY is a much more sensitive reporter than SpyA. Repetition of this simple experiment with a non-binding phosphorus compound (`MFCD00057611`) resulted in no significant change in linewidth.

Use of a ^{31}P-edited inversion-recovery pulse sequence revealed that the T_1 relaxation time constant for free phosphotyrosine is approximately 4.3 seconds. In order to obtain maximal signal intensity, a period of $5\times T_1$ seconds must occur between the final pulse of one scan and the first pulse of the next. In comparison to SpyA, whose ^{19}F T_1 was 0.72 s, this enforces a 6-fold increase in CPMG acquisition times. To counteract this situation, the 90°_x pulse width is reduced, causing a smaller angular rotation of the bulk magnetisation vector. This effectively decreases the time required to re-establish initial equilibrium spin orientations, however it has a negative impact on the overall signal intensity. The optimal balance between scan rate and available signal occurs at the 'Ernst angle'[‡]. A scan repetition time equal to T_1 was chosen, corresponding to an Ernst angle of 65 °. The duration of the first pulse in the CPMG sequence is therefore multiplied by 65/90, which in the case of phosphotyrosine generated a new pulse width of 5.05 μs. Under these conditions, the relative signal to noise is ~30 % lower than the maximum possible.

A proton-decoupled ^{31}P-edited version of the CPMG pulse sequence was used to analyse a sample containing 10 mM pY (spy) and `57611` (control). Over 200 CPMG loop cycles ($n2\tau = 8$ s) were required to remove signals from pY, indicating that its R_2 relaxation rate is much slower than that of SpyA, which disappeared from the equivalent ^{19}F-edited spectrum after 50 cycles (2 s). The experiment was repeated for 2 mM Spy and Control in the absence and presence of 100 μM *v*-Src SH2 using only 8 CPMG cycles. The domain caused a significant selective decrease in signal

[‡]Where complete decay of transverse magnetisation occurs between each scan, the optimal pulse tip angle for a given scan repetition time t_r is known as the Ernst angle, α_e according to $\alpha_e = e^{-t_r/T_1}$.[330]

intensity for the reporter ligand (data not shown). This confirms that interaction with SH2 leads to a dramatic increase in $R_{2,obs}$ for the Spy molecule.

Unfortunately, due to time limitations the further experimentation could not be performed; PAXS is a 'work-in-progress'. The encouraging results obtained with the SH2 domain system confirmed that a ^{31}P-based version of FAXS may be widely-applicable to other phosphate-binding targets, by virtue of the large chemical exchange contribution to observed R_2. Since the sensitivity of detection and the chemical exchange term are both proportional to field strength, the viability of PAXS will increase if high-field spectrometers and/or cryoprobes are available. Due to the relative insensitivity of the ^{31}P nucleus, high concentrations (200-300 μM) of reporter molecules are required. Test compounds may also be screened at this concentration, however weak-binding competitors will not be detected unless their relative concentration is raised. This may lead to problems with solubility and non-specific binding, and will reduce the size of mixtures that can be screened[334].

6.8 Summary

This chapter has described the practical application of FAXS to NMR-based fragment screening against the SH2 domain. The experiment was validated and optimised on a 400 MHz (9.4 T) spectrometer, which is considerably less sensitive than that employed by Dalvit *et al.* (600 MHz). Since chemical shift differences and the available signal intensities are proportional to field strength, the overall experimental sensitivity of FAXS in terms of detecting signal and the change in $R_{2,obs}$ is improved by moving to a higher field spectrometer. This reduces sample requirements and/or acquisition times, both of which increase screening throughput. Nevertheless, screening could be performed at 400 MHz as long as the quantity of receptor is not limiting, nor solubility of screening compounds a problem.

The SH2 domain system was found to be inferior in terms of screening efficiency compared to human serum albumin, the target used by Dalvit *et al.* This can be rationalised by the small size of the SH2 domain, coupled to the relatively low affinities it displays for fragment ligands (K_d >110 μM). Both these factors serve to weight the reporter ligand $R_{2,obs}$ towards that of the free state ($R_{2,f}$), which reduces difference in CPMG signal intensity between free and receptor-bound spy molecule. Since this is the fundamental observable in FAXS, the concentration of SH2 domain had to be raised significantly in order to shift $R_{2,obs}$ towards $R_{2,b}$. The improved screening efficiency associated with the Celltech target, Protein-X, can be explained by its larger size (200 kDa), its higher affinity for the reporter (IC_{50} = 10 μM), and the deep cavity in which the ligand binds, which is expected to provide a larger contribution from chemical exchange.

Comparison of SpyA, SpyB, and CT_{spy} reporter ligands indicated that the position of the ^{19}F nucleus is important in determining $R_{2,obs}$. In order to take advantage of the significant contribution from chemical exchange, the ^{19}F nucleus should be situated at or very close to the binding interface, and preferably in close proximity to a group that could induce a large chemical shift change, such as an aromatic ring or charged side chain. Optimisation of the position of the reporter nucleus should result in a decrease in sample requirements, however one must be aware that insertion of the ^{19}F nucleus (or nuclei) at the binding interface may have a negative impact on ligand affinity.

Whilst the broad concepts of FAXS are quite simple to understand, this research has highlighted some theoretical and practical difficulties associated with spy molecules that are in intermediate exchange with the receptor. Establishment of a 'standard curve' that relates the proportion of bound ligand to CPMG signal intensity requires knowledge of the chemical shift differences associated with binding and the mean lifetime of the bound state in addition to the transverse relaxation rates of the free and bound states. If this information is available, one can calculate a dissociation constant for a competitive ligand from a single experiment by virtue of its effect on the signal intensity of the spy molecule.

The validity of a ^{31}P-based version of FAXS was also explored. Interaction of phosphotyrosine with the SH2 domain resulted in a significant increase in linewidth, which was not observed in similar experiments with fluorinated spy molecules. It is highly probable that the formation of multiple hydrogen bonds with phosphate and neutralisation of local charges induces a large change in ^{31}P chemical shift, which appears to have a dramatic effect on $R_{2,obs}$. These initial findings suggest that PAXS may be highly suited to screening against targets known to interact with phosphate-based ligands. The inherently low signal to noise ratio experienced in ^{31}P-NMR studies is the largest obstacle to overcome in optimising PAXS for screening. A higher field strength and/or cryoprobe technology would assist in this matter. If this issue can be overcome then the ^{31}P- and ^{19}F-edited versions of this experiment should be equally powerful in the identification of competing ligands. At the very least, where a fluorinated ligand is not available, the phosphorus nucleus offers a viable alternative. Another potential problem for PAXS is in the screening of phosphate-based fragments, which would appear in the CPMG spectrum and may overlap with the spy molecule resonances. In this case, one would need to engineer some deviation in chemical shift for the phosphate-based spy.

The *in vitro* screening of fragments against the SH2 domain resulted in the identification of three novel ligands – SpyA, SpyC, and 36757– that are close to the natural pY fragment ligand in terms of their structure and affinity. These results

again highlight the extreme preference of the SH2 domain towards aromatic phosphates. The binding thermodynamics of 36757 were distinguished from the other ligands by a large negative enthalpy, however this was offset by an unfavourable contribution from entropy to the overall free energy of binding. NMR experiments also confirmed that fluorophosphate has weak affinity for the SH2 domain, however it was still effective as a Spy molecule despite its unfavourable signal characteristics. The use of these novel fragment ligands in drug design for the SH2 domain is discussed in the next chapter.

Chapter 7

Discussion

This chapter discusses further the observations and conclusions that can be drawn from this research as well as the contributions made towards the understanding of SH2 domain structure, function and drug design. Potential uses are suggested for the fragment ligands identified and characterised in Chapters 5 and 6. Finally, from practical experience and extensive comparisons within the literature, the effective use of NMR in the screening and analysis of fragment ligands is discussed.

7.1 Conclusions regarding apo *v*-Src SH2

The determination of the solution state structure of *v*-Src SH2 in the apo form is both an important and novel achievement that supports the more specific biophysical analyses carried out in this research. To date, all Src family SH2 domain structures that are free of peptide ligand have been elucidated in the presence of anionic buffers that display weak affinity for the phosphotyrosine binding site (see Table 3.1). This genuine apo structure of *v*-Src SH2 aligns quite closely with the phosphate-bound crystal structure of *v*-Src SH2 described previously (1SPR[101]). The most notable differences were at the N-terminal end of the βC strand, which forms part of the pY binding site, and in position of EF and BG loops, which combine to form the pY+3 pocket (see Section 3.7.1). Despite the fundamental differences between the two structure determinations, the most important being the absence or presence of phosphate, it appears that the overall structure of the *v*-Src SH2 domain is insensitive to phosphate binding. The phosphate-bound form of *v*-Src SH2 was investigated in the solution state by observing changes in chemical shift and dynamics for backbone amides (see Sections 4.1.1 & 4.2.4). The chemical shift data indicated that the effects of phosphate binding on backbone amides are limited to the pY pocket, and there is clear evidence for a direct hydrogen bond with E35. The dynamics study revealed that pY pocket residues are largely unaffected by phosphate binding,

however the N-terminal half of the BC loop displayed a slight decrease in mobility. Other parts of the domain did not indicate significant sensitivity to this small, anionic ligand.

The atomic structure of *v*-Src SH2 described in Chapter 3 is of high quality and enables detailed understanding of the apo state of an SH2 domain. The rms deviation of backbone atoms within the ensemble is low for regions of regular secondary structure (0.25 Å), which indicates that in the absence of ligands the domain structure remains well-defined. Loop regions displayed slightly higher rms deviations from the mean structure, however the NMR spectra do not indicate any slow or intermediate conformational exchange processes occurring in these regions. Data describing the backbone dynamics of *v*-Src SH2 in the apo state agree very well with the calculated atomic structure, and there is a broad inverse correlation between structural precision and motional freedom (see Figure 4.11). Comparison of the dynamic profile of the two binding pockets of the *v*-Src SH2 domain suggests that the pY pocket is significantly less mobile compared to the pY+3 pocket (Figure 4.10). This can be rationalised by the considering the contrary structures of the binding sites – the former displays a high prevalence of secondary structure whilst the latter is mainly formed by loop regions. These differences in structure and hence dynamics are significant and provide insight into the observed binding behaviour at each site (discussed below). The investigation into side chain ionisation states revealed that His βD4 (H58) has a pK_a value below 4.0 and is therefore fully deprotonated at pH 6.0 in both the apo and peptide-bound forms (see Section 4.3.5). This state appears to be induced by formation of a hydrogen bond with Arg βB5 (R32), coupled with relatively high local positive charge. This effectively restrains the side chain of Arg βB5 and pre-orders this crucial moiety to interact with ligands. Thus, the dynamics of the pY binding pocket are regulated at the level of the backbone and the side chain.

7.2 SH2 inhibitor drug design

7.2.1 Targeting the pY pocket

The data collected for the apo and ligand-bound forms of *v*-Src SH2 are consistent with a general emerging picture of SH2 domain binding behaviour. The combined evidence from structural, chemical shift, and dynamics studies indicate that the pY binding site is configured or 'hard-wired' to interact with pY and that it has only a limited capacity to adapt to alternative ligands. Modification of the pY epitope is imperative since it is metabolically unstable and carries multiple negative charges that impede cellular uptake. The electrostatic and hydrogen bonding interactions

that mediate the binding of pY to *v*-Src SH2 are clearly understood and, superficially at least, suggest that it ought to be relatively straightforward to identify a ligand capable of making similar interactions. However, the geometric requirements dictated by the internal structure and dynamics of the pY pocket are surprisingly selective towards pY-like ligands and have largely confounded attempts to develop high affinity therapeutic inhibitors.

Since the pY binding site is positively-charged and thus preferentially interacts with anionic ligands the twin aims of reducing the negative charge of pY and increasing affinity may be mutually incompatible[335]. Calculation of the phosphotyrosyl pK_a (Section 4.3.7) revealed a decrease from 6.0 to 4.9 on binding to the *v*-Src SH2 domain, which is consistent with a preference for the PO_4^{2-} form. Waksman *et al.* observed reduced peptide binding affinities for pH values below 7.5 and calculated a similar change in pK_a value for pY[193]. Additionally, a range of phosphonate-based pY analogues (compare 6–10 in Section 1.4.2.1) indicated that the pK_a of the phosphonate group has a strong influence over binding affinity. Substitutions that reduce the pK_a of the group led to improvements in IC_{50} value[178]. Conversely, pY mimetics that are devoid of charge (e.g. vicinal dicarbonyls, see **18** & **19**, Section 1.4.2.1) display large decreases in affinity. It is telling that the vast majority of current SH2 domain ligands feature negative charge at this position, despite the much-vaunted aim of abolishing it. All these studies suggest that negative charge and hydrogen-bonding capacity are very important aspects of pY binding, and thus may not be avoidable in SH2 domain inhibitors.

The adverse effect of negative charge on cellular uptake (and oral bioavailability in general) has been tackled previously through use of the pro-drug approach[336]. In this scheme, charged groups are temporarily masked by covalent attachment of a suitable protecting group. Upon entry into the cell, the pro-drug is converted into the active form by some enzymatic and/or spontaneous chemical process. Indeed, this has been attempted with SH2 domain inhibitors (see Section 1.4.2.4) and protection does facilitate membrane permeability[177, 197, 337]. However, in one case the rate of conversion of these compounds to the high affinity F_2Pmp-based* inhibitor ranged from negligible to moderate. There have been very few reports describing the systematic exploration of pro-drugs for SH2 domains. Given the recognised need to deal with the issue, the dearth of publications may reflect a long-standing assumption that such charge can be removed, rendering pro-drug technology superfluous to SH2 domain inhibitors.

The second problematic issue with pY is its susceptibility to phosphatases. Replacement of the phosphoester oxygen of pY often results in high stability, yet

*F_2Pmp: 4-phosphonodifluoromethylphenylalanine

reduced affinity. This is rationalised by a loss of hydrogen bonding potential and an increase in the pK_a of the phosphonate group. Unfortunately, the compound that is closest to pY in terms of affinity, F_2Pmp (see 10, Section 1.4.2.1), displayed unacceptable toxicity in cellular assays[179]. Unlike the charge issue, protection against degradation need not be incompatible with improving affinity. However, the reduced affinity of stable pY analogues may continue to be a feature that has to be endured and additional binding energy sought through novel interactions with adjacent sites. The recent reports concerning the nature and conservation of SH2 domain surfaces provide good starting points for extending the scope of current SH2 ligands to novel regions of space[99, 338]. Indeed, there are already some examples where addition of particular capping groups has enhanced peptide binding affinity[202, 203]. If successful, the exploitation of unique areas of SH2 domain surface should also result in a compound of improved specificity.

7.2.2 Targeting the pY+3 pocket

The pY+3 pocket exhibits several characteristics that contrast with the pY binding site, yet make it equally intractable in terms of drug design. A large variety of substructures have been employed to interact with this site, however they are always found in the context of a larger ligand which engages both pockets. Thermodynamic analysis of pYEE**X**-based peptides, where the **X** amino acid was was varied, indicated that diversity at the pY+3 site is well-tolerated in terms of affinity (see Table 1.5)[117]. The pYEEG peptide resulted in the largest decrease in affinity within the series, however this reduction was only ~20-fold in magnitude. This is surprisingly small given that the ligand side chain has effectively been removed from the pY+3 pocket. In fact, the pYGEI peptide resulted in a slightly worse affinity compared to pYEEG, which indicates that disruption of the pY+1 interaction is as (and possibly more) important to the affinity of peptide binding for *v*-Src SH2. A similar calorimetric study of a series of non-peptide SH2 domain inhibitors that differed at the pY+3 side chain indicated dissociation constants within an order of magnitude of each other (unpublished personal observations). Bradshaw *et al.* calculated that within the high affinity peptide pYEEI, the pY fragment provides ~60 % of the free energy of binding; the remainder is shared amongst the -EEI motif[106]. Other synthetic ligands have featured bulky, flexible, hydrophobic groups (e.g. linear and cyclic hydrocarbons) or the side chain chemistry of Ile, Val, or Leu[194, 200, 220, 221]. Finally, it is also interesting to note that there have been no reports of individual fragments binding to the pY+3 pocket. All these observations concerning the high degree of binding promiscuity at the pY+3 pocket are consistent with evidence from these studies, which indicate that this secondary pocket is inherently more

adaptable to ligand chemistry, by virtue of its dynamic, loop-based structure (see Section 4.2.3).

Modification of the pY side chain has been the major focus in the design of SH2 domain inhibitors; much less attention has been paid in the literature towards optimising interactions with the pY+3 binding site. Examination of the apo *v*-Src SH2 structure determined in these studies indicated the presence of several polar side chains in and around this secondary pocket that could be exploited to increase both the affinity and specificity of ligands. These are T72, R74, T75, Y87, and D92. Certain backbone amides of the EF and BG loops may also be targeted. It would be interesting to repeat the fragment docking approach employed in Chapter 5 to identify novel ligands for this site. These could be tested for binding using the $^{1}H^{15}N$-HSQC experiment, although high fragment concentrations (>5 mM) are likely to be necessary. Alternatively, such fragments could be conjugated to pY, or a suitable bioisotere, and screened for binding affinity. A similar process was recently applied to the Lck SH2 domain in an attempt to identify ligands that omit pY completely[339]. A database of 2 million compounds was filtered and screened *in silico* against the pY+3 pocket of Lck SH2 using DOCK. The highest-scoring 25000 small molecules were then docked again using a more rigorous method. These compounds were clustered according to similarity and this guided the final selection of 288 small molecules[†]. Of these, 34 compounds displayed significant inhibitory activity at 100 μM in an affinity precipitation assay, however only two ligands showed strong inhibition at lower concentrations. It is not clear at what sites these ligands interact with the Lck SH2 domain *in vitro*. It is possible that they bind to the pY pocket, although from a cursory analysis of the structures this does not appear to be likely. Attempting to circumvent the pharmacological issues associated with pY by ignoring its binding site in the design of ligands is an interesting and novel approach, which may result in a medium to high affinity specific ligand. However, one cannot ignore the possibility that such ligands may be conjugated to a pY mimetic fragment and achieve even higher affinity.

7.3 SH2 domains and *in vivo* signalling specificity

The SH2 domain continues to present a major challenge to the biological and pharmaceutical research communities. For such an apparently tractable and conceptually simple biological entity, there remains a number of perplexing issues that need to be overcome. For biochemists, the difficulty has not been the identification of SH2 domain binding sites, nor structural characterisation of the interface, but rather the

[†]Only 196 of these were acquired and tested *in vitro*.

obtaining of a sound understanding of SH2 domain binding specificity. Initially, it appeared likely that residues C-terminal to phosphotyrosine in target peptides would fulfil this role and this explanation was prematurely celebrated. It is a proven fact that isolated SH2 domains display a range of affinities towards phosphotyrosyl peptide sequences[110, 111] and that the preference of a given domain can be altered through targeted mutagenesis[107, 114, 115]. Thus it is logical to assign some level of specificity to this mechanism. However the difference in affinity between so-called 'specific' or 'high-affinity' peptides and their 'non-specific' counterparts is commonly just two orders of magnitude[119]. NMR relaxation analysis of *v*-Src SH2 indicated that the loop regions that define the pY+3 pocket are relatively mobile in the apo state, but undergo a decrease in mobility on interaction with PQpYEEIPI peptide. Insertion of the isoleucine side chain appears to stabilise these two loops regions. As described above, the dynamic behaviour of the pY+3 is consistent with the observed ability of SH2 domains to interact with other substructures and accounts for the relatively small differences in affinity between specific and less specific peptides.

It is not clear how this apparent lack of specificity observed for isolated SH2 domains and oligopeptides relates to *in vivo* signal transduction. Some studies of *in vivo* SH2 domain interactions have yielded findings that agree with the relatively low levels of specificity observed *in vitro*. For example, mutation of the Grb2 SH2 binding site on Met tyrosine kinase indicated that the two closely-spaced pY residues are absolutely necessary to embryonic viability[340]. However, substitution of the pY+2 site, which was also believed to be essential for specific Grb2 binding[341], only resulted in defective muscular development. Similarly, mutation of the SH2 domain of Grb2 homologue Sem-5 in *C. elegans* completely abrogated vulval development[114]. Microinjection of a Sem-5/SrcSH2 chimeric protein rescued the phenotype to approximately 30 % vulval development compared to wild type. Both of these examples indicate that specific recognition between SH2 domains and target sequences is not entirely necessary for effective SH2-mediated signalling.

Others have reported contrasting observations. For example, *in vivo* analysis of chimeric constructs of *c*-Fes kinase in which the SH2 domain was replaced by that of *v*-Src, *v*-Fps, or GAP indicated that indicated that only the Fes/Src construct was active[342]. However, this fusion protein exhibited potent transforming activity and induced developmental changes in fibroblasts and myeloid leukemia cells, respectively, which deviates from the normal cellular function of *c*-Fes.

Although current *in vivo* evidence is limited, it appears that individual SH2 domains are able to recognise a range of alternative phosphotyrosyl peptide sequences. This presents a major conceptual challenge regarding how specificity in SH2-mediated signalling is achieved. For propagation of such signals via linear

protein-protein interaction pathways it is clear that highly specific recognition is required[116]. Since the evidence for the SH2 domain does not appear to support this relatively simple model of eukaryotic signal transduction, the situation *in vivo* must be more complex than is currently understood. Indeed, the fact that many ligands designed to target specific signalling proteins display diverse toxic effects in cellular assays indicates that such compounds may be inhibiting a complex system rather than a simple linear cascade.

Over the past few decades it has become increasingly apparent that eukaryotic signalling proteins are generally assembled from a group of homologous interaction modules, such as the SH2 domain, and that specific multimeric complexes are formed in response to particular stimuli through these recognition domains[2, 116, 120]. The combined effects of several interactions may enhance the affinity of SH2 domain towards its ligands and provide an additional level of specificity. Additionally, some proteins (e.g. ZAP70) feature tandem SH2 domain arrangements, which may lead to increased affinity when interacting with a target bearing two phosphotyrosine sequences. However, recent studies have suggested that cooperativity between such domains does not necessarily result in improved ΔG_b° values, perhaps because of an additional entropic penalty incurred through the spatial ordering of the tandem domains[118, 234]. In conclusion, our current understanding of how much SH2 domain specificity is necessary or indeed how much actually exists *in vivo* is rather limited. The development of specific, high affinity SH2 inhibitors is driven by a therapeutic need, however such compounds will also be extremely useful in unravelling the complex web of molecular recognition events involved in eukaryotic signal transduction.

7.4 Fragment approach applied to the SH2 domain

The fragment-based approach employed in the search for novel ligands for the pY binding site was effective. The FLEXX docking software was able to predict the specific interaction with phosphotyrosine, and yielded highly favourable scores for a range of other known fragments. The docking results for the ~13000-member fragment library highlighted a number of interesting compounds that were found to be chemically similar to known ligands, and a very small group were purchased for analysis. The FAXS screening experiment was straightforward to optimise and enabled rapid identification of fragment ligands. The acquisition of thermodynamic data describing fragment ligands provided an additional and useful level of information, since ligands of similar affinity displayed variable enthalpic and entropic contributions to binding. The small library of fragment compounds will continue to be useful in the analysis of fragment screening experiments, which could employ techniques besides NMR.

Although the fragment-based approach has proved useful in the identification of pY mimetics, is may be less applicable to the pY+3 pocket since the binding energy available is much lower. Although the study of Huang *et al.* described above did target this secondary site, their ligands were not fragment-sized and hence are likely to contact additional regions of the SH2 domain surface. Moreover, addition of a pY mimetic to such ligands may result in unacceptable increases in molecular weight that inhibit lead development. Instead, the fragment-based approach for the SH2 domain may require gradual extension of a suitable pY analogue towards the pY+3 pocket and engaging multiple adjacent regions of the domain surface in a productive manner. This is a much slower, more design-intensive method of achieving a suitable therapeutic ligand compared to the conceptually simple case where two medium-affinity fragments are conjugated.

7.5 Use of novel pY mimetic ligands

This section describes the potential uses of SpyA, 36757 and fluorophosphate in the design of novel SH2 domain inhibitors. Since SpyA and 36757 (see Figure 7.1) contain labile phosphate groups they are vulnerable to degradation by phosphatases. Indeed, the non-fluorinated analogue of SpyA, 4-methyl-umbelliferylphosphate (MUP) and its derivatives are often used as a fluorogenic substrates for alkaline phosphatase and protein tyrosine phosphatases[343]. Thus some protection against metabolic degradation is required. Difluoromethylphosphonate derivatives of 36757 and SpyA (1 and 2, respectively) should improve stability and reduce the phosphonate pK_a value as close as possible to that of phosphate.

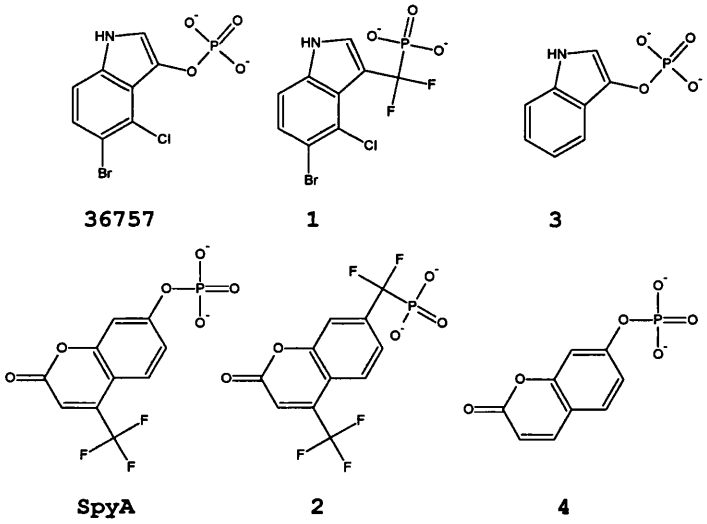


Figure 7.1 Possible changes that could be made to SpyA and 36757 to improve their metabolic stability and reduce unnecessary molecular weight.

The molecular weight of each fragment could be reduced by substituting heavy atoms that do not appear to be important to binding. For **36757**, removal of the Br and Cl atoms results in a substantial saving of 115 Da; the molecular weight of the indolylphosphate fragment (**3**) is just 209 Da. Similarly, for binding of SpyA, the CF_3 group may be superfluous, and its replacement with hydrogen (**4**) reduces molecular weight by 69 to a total of 239 Da. However, the removal of electron-withdrawing groups (Br, Cl, CF_3) from the aromatic ring systems of **36757** and SpyA may cause a decrease in binding affinity by modulating the pK_a of the phosphate group. Changes in affinity can be measured using competition ITC experiments, and the pK_a calculated through analysis of ^{31}P-NMR chemical shifts. A series of fluorine ring-substitutions could then be explored in an attempt to recover the lost affinity. Fluorine is strongly electronegative, and it has the lowest mass of the halogen elements (18 Da).

The ultimate usefulness of **36757** and SpyA and their derivatives will be determined by their ability to mimic phosphotyrosine within the context of a larger scaffold. To investigate this, peptide analogues of pYEEI could be synthesised in which the phenylphosphate side chain is replaced by these fragments. Since the bicyclic ring system of **36757** and SpyA make them larger than pY, a number of different structures may need to be synthesised in order to optimise ligand affinity. In the natural pYEEI ligand, the amide of the first glutamate forms a hydrogen bond with HisβD4 backbone carbonyl in the SH2 domain. This acceptor was also implicated in the binding of **36757** (see Figure 6.25). It is probable that the former interaction would dominate *in vitro* since it requires neither the hypothesised deformation of the βD strand, nor the presence of a bridging water molecule. The loss of this hydrogen bond in binding **36757** may reduce its usefulness as a pY mimetic.

As described in Section 1.4, the instability of peptidic SH2 inhibitors prevents their use as efficient therapeutics. Instead, **36757** and SpyA will need to be attached to non-peptide scaffolds that are associated with higher stability, cell penetration, and specificity. For *c*-Src SH2, the latest generation of inhibitors fall into two classes, defined by AP22408 and RU81843, which were developed between 2000 and 2002 by *ARIAD*[159] and *Aventis*[217, 218], respectively (see Figure 7.2). These compounds were 15–20-fold more potent than pYEEI in binding assays. Possible replacements of the bisphosphonate group of AP22408 (**5** & **6**) and the phosphotyrosyl group of RU81843 (**7** & **8**) with either **36757** or SpyA are shown in Figure 7.2. Note that the point of attachment for each fragment should be varied in order to determine the optimal configuration for binding.

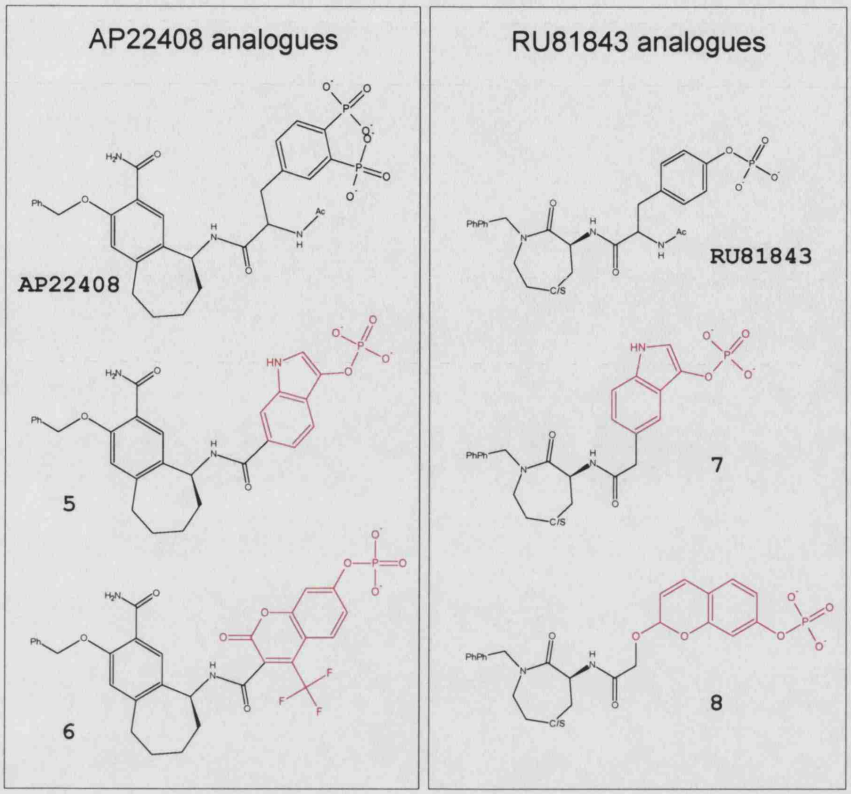


Figure 7.2 Use of 36757 and SpyA to replace phosphotyrosine in AP22408 and RU81843, which represent two of the most advanced Src SH2 'high-affinity' scaffolds.

7.5.1 Fluorophosphate replacements

Fluorophosphate is another interesting fragment ligand. Like phosphate, this compound was found to bind to *v*-Src SH2 with weak affinity and it is probable that fluorophosphotyrosine (9, Figure 7.3) would display a similar affinity to pY. The deviation from phosphate could render 9 less susceptible or even completely resistant to protein tyrosine phosphatase-mediated hydrolysis. Interestingly, Venetz *et al.* report the hydrolytic cleavage of the P–F bond of monofluorophosphate by alkaline phosphatase[344]. Incorporation of FPO_3 into uridine 5'-phosphate conferred resistance to this enzyme, however it was still sensitive to 5'-nucleotidase, which catalyses the hydrolysis of phosphate from nucleotides[345]. The rate of conversion of uridine 5'-fluorophosphate to uridine was somewhat slower than that observed for natural phosphate substrate indicating that FPO_3 replacements may well be more resistant to phosphatase-mediated hydrolysis than PO_4-based analogues.

When used to modify either SpyA or 36757 (10 and 11), fluorophosphate could provide an attractive phosphate mimetic that has both reduced charge and is stable. A pro-drug derivative (e.g. 12) could mask the negative charge and enable

more efficient cellular uptake. In this regard, fluorophosphate compounds have had a chequered history. They have been used in everything from toothpaste (Na_2FPO_3) to fertilisers (organophosphates) to deadly nerve agents in chemical warfare (diisopropylfluorophosphate (DFP), 13). Although this last compound is a highly potent cholinesterase inhibitor, it is currently employed as a protease inhibitor in biological research. There exists some chemical similarity between 12 and DFP however it is hoped that this pro-drug derivative of fluorophosphotyrosine will not share in its toxic nor anti-protease effects. Other pro-drug forms may be more efficient than those indicated in Figure 7.3.

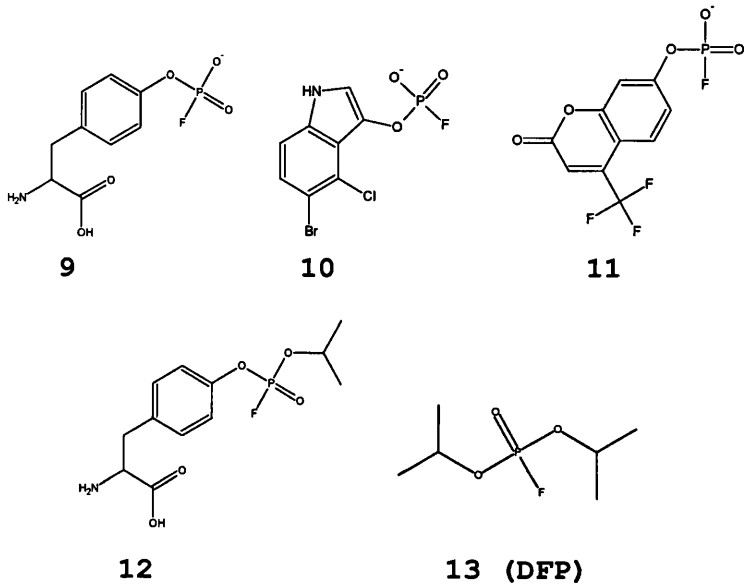


Figure 7.3 Use of fluorophosphate to replace phosphate in pY, 36757, and SpyA.

7.5.2 Novel fluorescence-based approaches for studying SH2 domains

As mentioned above, the SpyA analogue MUP (Figure 7.4) is a fluorogenic substrate for alkaline phosphatase[343]. The difluorinated derivative DiFMUP is also exploited in the analysis of protein Tyr and Ser/Thr phosphatase activity[343]. Whilst the discovery that a SpyA interacts with *v*-Src SH2 is novel and may yield a useful pY mimetic fragment, the fact that it is a fluorophore may enable novel applications of fluorescence towards the study of SH2 domain interactions and/or screening for ligands.

Due to the relatively high sensitivity, simplicity, rapidity and low cost of technology, fluorescence is gradually becoming the technique of choice in HTS programs[346, 347]. Lynch *et al.* (*ARIAD*) developed a fluorescence polarisation-based assay for measuring the potency of *c*-Src SH2 domain ligands[171]. In this experiment, competition

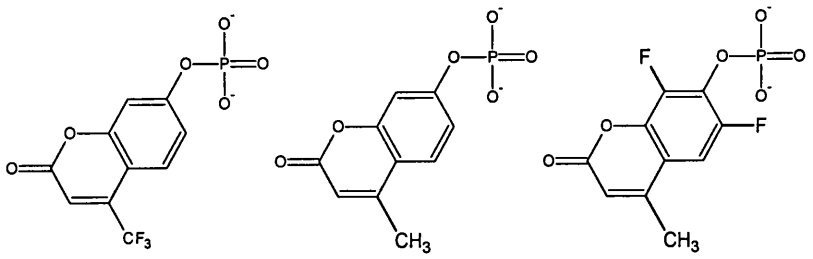


Figure 7.4 Fluorogenic SH2 domain ligands

between a fluorescein-tagged peptide (f-GpYEEIA-NH_2) and another ligand affects the polarisation behaviour of the fluorophore[277, 348]. The discovery that a fragment-sized fluorophore binds directly to the SH2 domain immediately suggested the use of fluorescence spectroscopy as another fragment screening technique. This is not found in the literature, probably because there are very few fluorophores available that are known to bind to specific protein sites of interest.

This novel possibility raises several key questions. Firstly, can fluorescence techniques be used to detect the binding of MUP derivatives, such as SpyA, to the *v*-Src SH2 domain? Secondly, how do the sample requirements and overall experimental sensitivity compare to NMR, and other fragment screening approaches? Finally, what experimental conditions are necessary to detect competition between the fluorophore ligand and a competitor, such as pY? If favourable answers to these questions can be found then one could feasibly perform a fluorescence-based screen of thousands of compounds using a multiwell plate in a few minutes. Of course, fluorescence is not able to identify the binding ligand from a mixture, nor provide structural information. However, this method could be used to pre-screen a large library of fragment compounds prior to analysis of the hits by NMR or x-ray crystallography and could result in a substantial saving of time and resources.

The major drawback of a competition-style fluorescence screening approach is the requirement for the reporter ligand to act as a fluorophore. This is not expected to be a common occurrence. However, given the fact that all SH2 and PTB domains bind pY, they will probably all bind fragments like MUP. This enables screening against a few hundred potential target already. Protein tyrosine phosphatase domains are the next obvious candidates, however a non-hydrolysable analogue of MUP will be required. Pastula *et al.* developed an assay kit for quantification of Ser/Thr protein phosphatase activity that is based on hydrolysis of DiFMUP, so clearly these enzymes are also capable of binding these fluorogenic fragments[349]. Thus application of fluorescence to fragment screening may be possible against a select range of targets that interact with phosphorylated peptides.

Unfortunately, SpyA, MUP, and DiFMUP are much more fluorescent in the non-phosphorylated form[‡]. However, given that MUP derivatives undergo enzymatic hydrolysis by phosphatases, SpyA may be utilised to study the weak phosphatase activity of the Src SH2 domain and explore the potential role of Cys βC3. The pH titration data described in Chapter 4 indicated that Cys βC3 of the *v*-Src SH2 domain displays a pK_a value that is slightly below-average. This accounts for the small decrease in peptide binding affinity observed at pH values above 7.5 and for its reactivity as a nucleophile[193]. This unique feature has been exploited in the past to design Src-specific SH2 inhibitors through covalent attachment approaches[160, 194]. Cys βC3 has also been implicated in the very weak phosphatase activity of this SH2 domain[298]. Solutions containing pYEEI or *p*-nitrophenylphosphate revealed slow hydrolysis and formation of inorganic phosphate. It is possible that Src SH2 will also catalyse the degradation of SpyA or DiFMUP, and this could be monitored easily by detecting the fluorescence of the reaction product. It is still not clear why Cys βC3 is conserved only in Src SH2 domains and what functional role it plays, if any, *in vivo*. Thus SpyA analogues could be useful in resolving this enigma.

7.6 NMR screening strategies

The use of NMR in the screening of compounds continues to increase in scope as novel experimental approaches are developed. At present, there are five alternative pulse sequence types[§] that include 29 potential screening experiments. In addition to this, screening can also be performed in competition mode, yielding up to 58 combinations. Several of the novel combinations available (e.g. spin-label enhanced FAXS) have not been proposed in the literature. With this large array of possible approaches comes a need to rationalise the selection of experiments in order to optimise screening efficiency and accuracy and provide the required information. A possible scheme, developed from a critical comparison of the available techniques and practical experience with FAXS and HSQC, is described below.

The choice of NMR screening technique is strongly influenced by the molecular weight of the receptor (Figure 7.5). Screening against low molecular weight targets (e.g. oligonucleotides, oligosaccharides, and oligopeptides) is difficult since the change in ligand rotational (τ_c) or transverse diffusion coefficients on binding are not easily detectable. Although this situation excludes the majority of approaches, NMR screening may still be performed by spin-labelling or immobilising the receptor, which artificially raises $R_{2,b}$, or instead by observing chemical shifts for

[‡]Source: *Molecular Probes Inc.* website (http://probes.invitrogen.com)

[§]These are: R_1-edited (realistically, competition-style only), R_2-edited, diffusion-edited, NOE-edited, and chemical shift-edited

the receptor. At the other extreme, when one considers membrane-associated, or otherwise immobilised receptors, very few techniques remain applicable. Only two methods have been used to screen against such targets – saturation transfer difference (STD)[82, 83] and target immobilised NMR screening (TINS)[77]. For intermediate molecular weights, there is greater freedom of choice, however diffusion-based techniques (including NOE pumping) have an upper limit of approximately 100 kDa¶. The range of applicability for target-based screening (HSQC) is particularly dependent on the available sample and equipment. For relatively small receptors (10-35 kDa), a uniformly ^{15}N-labelled sample will suffice. For larger targets, sensitivity must be increased by using selective labelling strategies and state-of-the-art spectrometers (with cryoprobe). This extends the range of HSQC (or TROSY) screening experiments up to 100 kDa.

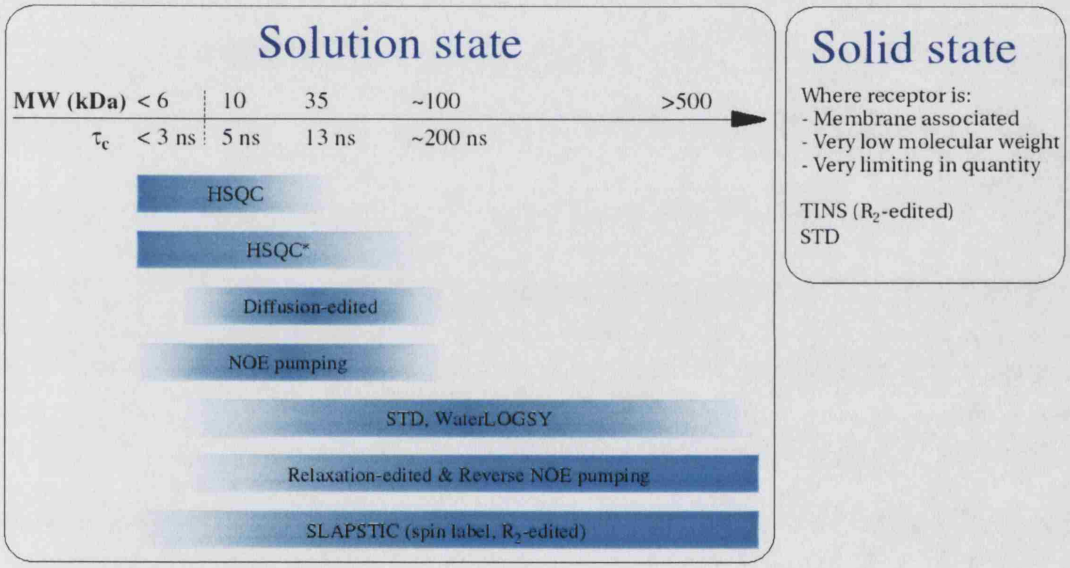


Figure 7.5 The approximate range of receptor molecular weights and τ_c values acceptable for NMR screening experiments in the solution and solid states. The viability of each technique is reduced at certain extremes; this is indicated by the colour gradients. HSQC* represents the target-based chemical shift method, but using selective isotopic labelling, the highest sensitivity equipment available, and possibly a TROSY-type pulse sequence instead.

The ligand dissociation constants anticipated during NMR screening will determine whether competition-style experiments are necessary. The threshold is reached at approximately $K_d \approx 0.1$–1.0 μM; below this value, false-negatives will arise during standard screening. To date, competition NMR experiments have not been reported for diffusion-based approaches or for reverse NOE pumping although, in principle,

¶Interaction between a ligand and a large receptor results in a rapid loss of phase coherence (large R_2) for the ligand, which is exploited in the relaxation-edited experiment, but leads to false-negatives in the diffusion-edited scheme in which high protein:ligand concentration ratios are optimal.

these are possible. The advantageous features of competitive screening, mainly reduced spectral complexity and access to ligand dissociation constants, encourages its use even when it is not strictly necessary. This was certainly true for the application of FAXS during this research. However, the efficiency of competition-style screening will tend to decrease as the reporter ligand K_d approaches mM values since the required concentration of receptor and test compounds will increase.

Competition-style experiments are not necessary in target-based screening as there is no lower limit to K_d and there is rarely any practical advantage in using a reporter ligand. Moreover, such methods can be optimised for selective detection of high affinity ligands by lowering the concentration of test compounds, hence there is no need to employ a known ligand to compete out weak-affinity compounds. Consequently there have been no reports in the literature concerning competition-based HSQC screening. Nevertheless, where the stability or folding status of a protein is dependent on the presence of a particular ligand, target- or ligand-based NMR screening performed against this complex may involve competitive equilibria. Depending on the binding site of interest, the presence of the essential ligand may inhibit the detection of weak-binding competitors, unless their concentration can be raised sufficiently.

The next stage in experimental selection excludes approaches that are inaccessible due to practical limitations. A low quantity of receptor sample will favour those techniques that have minimal requirements (e.g. STD, relaxation-edited). Target can be reused if it is immobilised to a semi-solid support (e.g. sepharose[77] or biomembrane[83]). Other limitations are more obvious: lack of correct probe or solid state technology; lack of reporter ligand; inability to spin-label the receptor effectively.

The final aspect to consider in the selection of a suitable NMR screening experiment is the type of information required – to access certain data, one must perform the correct experiment. The information available from each technique is summarised in Table 1.3 on page 58. For medium- to high-throughput screening protocols, the power to identify a ligand directly within a mixture of non-binding compounds without the need for deconvolution experiments may dominate the choice of approach. Of course, this is incompatible with competition-style experiments. Conversely, for smaller screens where structural data (binding site and epitope) is preferred, NOE- or target-based approaches are more suitable. Non-competitive approaches can generally be used to obtain approximate affinity data. A more robust 'single-sample' measurement can be achieved using a reporter ligand. A standard curve may be created prior to screening, or predicted on the basis of theoretical equations.

To summarise, selection of the optimal NMR screening experiment for each situation should be guided first by the system of interest (molecular weight of receptor & anticipated ligand K_d), then by practical considerations (sample availability and technology), and finally by the type of information desired. In many cases, the most efficient experiment may not be the one that yields the required information. Therefore, the selection order could be reversed so that one first identifies experiments that will generate the appropriate data, and then filters the list according to the physical parameters of the system and any practical issues.

There is no theoretical reason why two or more NMR screening experiments cannot be performed on each sample in order to generate a complete set of information. As long as the concentration regimes required for each experiment are roughly overlapping, suitable approaches can be run consecutively. One of the most efficient, widely-applicable, and powerful combinations should be FAXS–STD. In the first experiment, the fluorinated reporter ligand is observed in the presence of test compounds, and if a competing ligand is present, there will be a measurable change in the ^{19}F-NMR spectrum. In the following ^{1}H-edited STD experiment, the reporter ligand will of course be identified, however signals from the competitor will also appear in the difference spectrum. Therefore, with this combination of approaches, one can identify the competitor from within the mixture, measure its affinity, verify its binding site[‖], and even determine the binding epitope on the basis of STD signal patterns. There is no single approach that can provide all this information. These data can be recorded using just one sample, of relatively low concentration, in about 12-20 minutes. If competing ligands are absent from the sample then no change will occur in the FAXS spectrum and the STD experiment will not need to be run. Since protons are selectively observed in the second experiment, mixtures will have to be designed carefully to avoid resonance overlap between test compounds. To date, there have been no reports in the literature describing concatenation of NMR screening experiments in this manner.

The future of NMR in the screening of fragment ligands, and in enabling fragment-based drug discovery in general, looks bright. Great progress continues to be made in the design of experiments and hardware to increase the efficiency of screening. Whilst this is a worthwhile task, care must be taken to avoid a particular failing of HTS where the apparent usefulness of a screen is measured by sheer numbers of compounds and not by the quality of the hits. NMR screening is capable of operating at HTS rates; this technology should arrive within the next decade. However the aim of the fragment-based approach to drug design is not to screen as many

[‖]By virtue of the FAXS competition experiment, the ligand either competes directly with the Spy, or it may bind elsewhere and cause a conformational change that displaces the Spy. Either way, the ligand is potentially interesting.

fragments as possible, but rather to make intelligent use of weak-binding fragment ligands, extending and linking them productively to generate a high-affinity ligand. As long as the pharmaceutical industry remains close to this intellectually more difficult task, the fragment approach will be more successful in developing novel therapeutics than existing high-throughput methods.

Appendix A

Preparation of minimal medium

This section describes the method used to preparing one litre of minimal medium with any combination of ^{13}C and ^{15}N labelling.

6 g Na_2HPO_4
3 g KH_2PO_4
0.5 g NaCl
1 g $(NH_4)_2SO_4$ – substitute ^{15}N-labelled ammonium sulphate as required

The above compounds are added to 1 litre ddH_2O, the pH adjusted to 7.4 and the solution autoclaved. After the solution has cooled to room temperature, the following are filter sterilised and added:

2 ml of 1 M $MgCl_2$
100 μL of 0.1 M $CaCl_2$
100 μL of 0.1 M $FeSO_4$ (or $FeCl_2$) freshly made
2 g glucose – substitute ^{13}C-labelled glucose as required
1 mL of 1x micronutrient solution
1 mL of 1x vitamin solution

Micronutrient 1000x stock solution recipe, stored at -20 °C

0.0726 g Na_2MoO_4
2.473 g H_3BO_3
0.713 g $CoCl_2 \cdot 6H_2O$
0.25 g $CuSO_4 \cdot 5H_2O$
1.853 g $MnCl_2 \cdot 4H_2O$
0.2876 g $ZnSO_4$
in 100 mL ddH_2O

Vitamin 1000x stock solution recipe, stored at -20 °C

0.04 g Acetylcholine chloride
0.05 g Folic acid
0.05 g Pantothenic acid
0.05 g Nicotinamide
0.10 g Myo-inositol
0.05 g Pyridoxal HCl
0.05 g Thiamine HCl
0.005 g Riboflavin
0.1 g Biotin
in 100 mL deionised distilled water

Appendix B

Coupling constant data

$^3J_{HNH\alpha}$ coupling constant data were calculated from empirical crosspeak volumes using 3.2. The A, B and C Karplus coefficients for fitting were set to 6.4, -1.4 and 1.9, respectively, whilst the phase was -60°. During structure calculations, all coupling constants were given error bounds of ± 0.5 Hz.

Residue	$^3J_{HNHA}$ (Hz)	Residue	$^3J_{HNHA}$ (Hz)	Residue	$^3J_{HNHA}$ (Hz)
2	6.38	38	4.54	74	6.80
3	7.15	40	9.98	75	9.95
4	3.52	41	9.21	76	10.17
5	7.78	42	9.97	77	10.70
6	8.87	43	9.74	78	6.24
7	9.86	44	10.14	79	9.88
9	8.04	45	10.69	80	2.97
10	6.99	46	7.41	81	3.42
11	6.03	47	8.61	82	5.54
12	2.78	48	8.56	83	4.39
13	3.16	49	7.62	84	3.39
14	6.31	50	3.67	85	3.97
15	3.61	51	9.31	86	5.30
16	6.10	52	8.89	87	7.35
17	4.19	54	4.92	88	6.97
18	5.34	55	8.15	89	6.84
19	4.54	56	9.10	90	9.63
20	7.43	57	10.63	91	5.14
21	4.56	58	9.15	92	6.14
23	6.33	59	9.74	94	7.76
24	5.55	60	6.50	95	3.69
26	3.70	61	9.83	96	6.67
28	6.10	62	9.33	97	5.10
29	12.40	63	9.61	98	6.60
30	6.31	64	8.34	99	10.73
31	10.74	65	3.66	100	7.80
32	9.73	66	7.40	101	6.60
33	7.35	69	9.53	102	1.88
34	4.61	70	7.69	104	8.60
35	6.59	71	11.21	105	8.64
36	10.46	72	10.04	106	9.05
37	8.70	73	3.55		

Table B.1 $^3J_{HNH\alpha}$ coupling constants for *v*-Src SH2.

Appendix C

Temperature coefficient data

Amide proton temperature coefficients were calculated for *v*-Src SH2 as described in Section 3.9.3.

Residue	dδ/dT (ppb/K)	Residue	dδ/dT (ppb/K)	Residue	dδ/dT (ppb/K)
2	-7.458	38	-7.727	72	-6.057
3	-8.100	39	-4.083	73	-8.104
4	-6.277	40	-1.458	74	-4.708
5	-1.075	41	-4.449	75	-2.667
6	-2.478	42	-2.042	76	-7.225
7	-3.005	43	-2.523	77	-2.903
8	-0.375	44	-2.792	78	-4.708
9	-7.607	45	-2.863	79	-1.792
10	-1.825	46	-7.625	80	-4.629
11	-3.500	47	-1.126	81	-6.417
12	-2.707	48	-1.958	82	-4.285
13	-5.792	49	5.375	83	-2.583
14	-7.030	50	-5.750	84	-4.092
15	-3.500	51	-2.612	85	-4.074
16	-1.040	52	-5.792	86	0.245
17	-3.750	53	0.000	87	-2.308
18	-4.708	54	-8.068	88	-0.375
19	-5.945	55	1.167	89	-1.830
20	-5.958	56	-7.042	90	-0.875
21	3.958	57	-3.494	91	-7.444
23	-4.125	58	-6.567	92	-6.083
24	-1.091	59	-3.667	93	-5.417
26	-6.750	60	-5.556	94	-4.632
27	-0.083	61	-1.644	95	-5.125
28	-3.381	62	-3.086	96	-1.333
29	-1.247	63	-7.250	97	-9.468
30	-2.167	64	-4.912	98	-3.381
31	-3.875	65	-8.732	99	-3.333
32	-2.416	66	-4.250	100	-4.661
33	-2.965	67	-2.667	101	-5.398
34	-2.375	68	-4.208	102	-5.042
35	-5.633	69	-3.417	104	-6.917
36	-2.866	70	-1.827	105	-7.833
37	-4.700	71	-4.706	106	-8.750

Table C.1 Temperature coefficients ($d\delta/dT$) for *v*-Src SH2 in units of ppb/K.

Bibliography

[1] PAWSON T & NASH P (**2003**). Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science*, 300, 445–452.

[2] MACHIDA K & MAYER BJ (**2005**). The SH2 domain: versatile signalling module and pharmaceutical target. *Biochemica et Biophysica Acta*, 1747, 1–25.

[3] EDITORIAL (**2005**). Drug safety on trial. *Nature*, 434, 545.

[4] WADMAN M (**2005**). Drug safety special: The safety catch. *Nature*, 434, 554–556.

[5] HIRD N (**2000**). Isn't combinatorial chemistry just chemistry? *Drug Discovery Today*, 5 (8), 307–308.

[6] MURRAY CW & VERDONK ML (**2002**). The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *Journal of Computer-Aided Molecular Design*, 16, 741–753.

[7] LAHANA R (**1999**). How many leads from HTS? *Drug Discovery Today*, 4 (10), 447–448.

[8] CAMPBELL S (**2000**). Science, art and drug discovery: a personal perspective. *Clinical Science*, 99, 255–260.

[9] HANN MM, LEACH AR, & HARPER G (**2001**). Molecular complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *Journal of Chemical Information and Computer Sciences*, 41, 856–864.

[10] JHOTI H (**2005**). A new school for screening. *Nature Biotechnology*, 23 (2), 184–186.

[11] BOHACEK RS, MCMARTIN C, & GUIDA WC (**1996**). The art and practice of structure-based drug design: A molecular modelling perspective. *Medicinal Research Reviews*, 16, 3–50.

[12] ERLANSON DA, MCDOWELL RS, & O'BRIEN T (**2004**). Fragment-Based Drug Discovery. *Journal of Medicinal Chemistry*, 47 (14), 3463–3482.

[13] GIBBS JW (**1878**). On the Equilibrium of Heterogeneous Substances.

[14] PAGE MI & JENCKS WP (**1971**). Entropic contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proceedings of the National Academy of Sciences U.S.A.*, 68 (8), 1678–1683.

[15] JENCKS WP (**1981**). On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences U.S.A.*, 78 (7), 4046–4050.

[16] HAJDUK PJ, SHEPPARD G, NETTESHEIM DG, OLEJNICZAK ET, SHUKER SB, MEADOWS RP, STEINMAN DH, CARRERA GM JR, MARCOTTE PA, SEVERIN J, WALTER K, SMITH H, GUBBINS E, SIMMER R, HOLZMAN TF, MORGAN DW, DAVIDSEN SK, SUMMERS JB, & FESIK SW (**1997**). Discovery of Potent Nonpeptide Inhibitors of Stromelysin Using SAR by NMR. *Journal of the American Chemical Society*, 119 (25), 5818–5827.

[17] GREEN N (**1975**). Avidin. *Advances in Protein Chemistry*, 29, 85–133.

[18] REES DC, CONGREVE M, MURRAY CW, & CARR R (**2004**). Fragment-based lead discovery. *Nature Reviews Drug Discovery*, 3, 660–672.

[19] WENDT M, ROCKWAY T, GEYER A, MCCLELLAN W, WEITZBERG M, ZHAO X, MANTEI R, NIENABER V, STEWART K, KLINGHOFER V, & GIRANDA V (**2004**). Identification of novel binding interactions in the development of potent, selective 2-naphthamidine inhibitors of urokinase. Synthesis, structural analysis, and SAR of N-phenyl amide 6-substitution. *Journal of Medicinal Chemistry*, 47 (2), 303–324.

[20] HAJDUK P, DINGES J, SCHKERYANTZ J, JANOWICK D, KAMINSKI M, TUFANO M, AUGERI D, PETROS A, NIENABER V, ZHONG P, HAMMOND R, COEN M, BEUTEL B, KATZ L, & FESIK S (**1999**). Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *Journal of Medicinal Chemistry*, 42 (19), 3852–3859.

[21] LADBURY J (**1996**). Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chemical Biology*, 3 (12), 973–980.

[22] CHUNG E, HENRIQUES D, RENZONI D, ZVELEBIL M, BRADSHAW JM, WAKSMAN G, ROBINSON CV, & LADBURY JE (**1998**). Mass spectrometric and thermodynamic studies reveal the role of water molecules in complexes formed between SH2 domains and tyrosyl phosphopeptides. *Structure with Folding & Design*, 6 (9), 1141–1151.

[23] SUNNERHAGEN M, DENISOV V, VENU K, BONVIN A, CAREY J, HALLE B, & OTTING G (**1998**). Water molecules in DNA recognition I: hydration lifetimes of trp operator DNA in solution measured by NMR spectroscopy. *Journal of Molecular Biology*, 282 (4), 847–858.

[24] SZCZEPANKIEWICZ B, LIU G, HAJDUK P, ABAD-ZAPATERO C, PEI Z, XIN Z, LUBBEN T, TREVILLYAN J, STASHKO M, BALLARON S, LIANG H, HUANG F, HUTCHINS C, FESIK S, & JIROUSEK M (**2003**). Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *Journal of the American Chemical Society*, 125 (14), 4087–4096.

[25] LIU G, XIN Z, PEI Z, HAJDUK P, ABAD-ZAPATERO C, HUTCHINS C, ZHAO H, LUBBEN T, BALLARON S, HAASCH D, KASZUBSKA W, RONDINONE C, TREVILLYAN J, & JIROUSEK M (**2003**). Fragment screening and assembly: a highly efficient approach to a selective, cell active protein tyrosine phosphatase 1B inhibitor. *Journal of Medicinal Chemistry*, 46 (20), 4232–4235.

[26] HUC I & LEHN JM (**1997**). Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly. *Proceedings of the National Academy of Sciences U.S.A.*, 94, 2106–2110.

[27] CHEESEMAN JD, CORBETT AD, GLEASON JL, & KAZLAUSKAS RJ (**2005**). Receptor-Assisted Combinatorial Chemistry: Thermodynamics and Kinetics in Drug Discovery. *Chemistry – A European Journal*, 11, 1708–1716.

[28] LEHN JM (**1999**). Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chemistry - A European Journal*, 5 (9), 2455–2463.

[29] RAMSTRÖM O & LEHN JM (**2000**). *In situ* generation and screening of a dynamic combinatorial carbohydrate library against concanavalin A. *ChemBioChem*, 1, 41–48.

[30] NGUYEN R & HUC I (**2001**). Using an enzyme's active site to template inhibitors. *Angewandte Chemie International Edition*, 40, 1774–1776.

[31] HOCHFURTEL M, KROTH H, PIECHA D, HOFMANN MW, NICOLAU C, KRAUSE S, SCHAAF O, SONNENMOSER G, & ELISEEV AV (**2002**). Target-induced formation of neuraminidase inhibitors from *in vitro* virtual combinatorial libraries. *Proceedings of the National Academy of Sciences U.S.A.*, 99, 3382–3387.

[32] PANG YP, QUIRAM P, JELACIC T, HONG F, & BRIMIJOIN S (**1996**). Highly potent selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase. Steps toward novel drugs for treated Alzheimer's disease. *Journal of Biological Chemistry*, 271, 23646–23649.

[33] ERLANSON DA, LAM JW, WIESMANN C, LUONG TN, SIMMONS RL, DELANO WL, CHOONG IC, BURDETT MT, FLANAGAN WM, LEE D, GORDON EM, & O'BRIEN T (**2003**). In situ assembly of enzyme inhibitors using extended tethering. *Nature Biotechnology*, 21, 308–314.

[34] LIEBESCHUETZ J, JONES S, MORGAN P, MURRAY C, RIMMER A, ROSCOE J, WASZKOWYCZ B, WELSH P, WYLIE W, YOUNG S, MARTIN H, MAHLER J, BRADY L, & WILKINSON K (**2002**). PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *Journal of Medicinal Chemistry*, 45 (6), 1221–1232.

[35] CARD GL, BLASDEL L, ENGLAND BP, ZHANG C, SUZUKI Y, GILLETTE S, FONG D, IBRAHIM PN, ARTIS DR, BOLLAG G, MILBURN MV, KIM SH, SCHLESSINGER J, & ZHANG KYJ (**2005**). A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nature Biotechnology*, 23 (2), 201–207.

[36] HOPKINS L, GROOM C, & ALEX A (**2004**). Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today*, 9, 430–431.

[37] ERLANSON DA & HANSEN S (**2004**). Making drugs on proteins: site-directed ligand discovery for fragment-based lead assembly. *Current Opinion in Chemical Biology*, 8, 399–406.

[38] ERICKSON JA, JALAIE M, ROBERTSON DH, LEWIS RA, & VIETH M (**2004**). Lessons in molecular recognition: The effects of ligand and protein flexibility on docking accuracy. *Journal of Medicinal Chemistry*, 47, 45–55.

[39] SHUKER SB, HAJDUK PJ, MEADOWS RP, & FESIK SW (**1996**). Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science*, 274 (5292), 1531–1534.

[40] BEMIS GW & MURCKO MA (**1996**). The properties of known drugs. 1. Molecular Frameworks. *Journal of Medicinal Chemistry*, 39, 2887–2893.

[41] BEMIS GW & MURCKO MA (**1999**). The properties of known drugs. 2. Side Chains. *Journal of Medicinal Chemistry*, 42, 5095–5099.

[42] FEJZO J, LEPRE CA, PENG JW, BEMIS GW, AJAY, MURCKO MA, & MOORE JM (**1999**). The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chemistry & Biology*, 6 (10), 755–769.

[43] FEJZO J, LEPRE C, & XIE X (**2003**). Application of NMR screening in drug discovery. *Current Topics in Medicinal Chemistry*, 3 (1), 81–97.

[44] MOORE JM (**1999**). NMR Techniques for Characterisation of Ligand Binding: Utility for Lead Generation and Optimisation in Drug Discovery. *Biopolymers*, 51, 221–243.

[45] MOORE J, ABDUL-MANAN N, FEJZO J, JACOBS M, LEPRE C, PENG J, & XIE X (**2004**). Leveraging structural approaches: applications of NMR-based screening and X-ray crystallography for inhibitor design. *Journal of Synchrotron Radiation*, 11 (1), 97–100.

[46] LEWELL XQ, JUDD DB, WATSON SP, & HANN MM (**1998**). RECAP-Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *Journal of Chemical Information and Computer Sciences*, 38 (3), 511–522.

[47] HAJDUK PJ, BURES M, PRAESTGAARD J, & FESIK SW (**2000**). Priveleged molecules for protein binding identified from NMR-based screening. *Journal of Medicinal Chemistry*, 43, 3443–3447.

[48] FATTORI D (**2004**). Molecular recognition: the fragment approach in lead generation. *Drug Discovery Today*, 9 (5), 229–237.

[49] LIPINSKI CA, LOMBARDO F, DOMINY BW, & FEENEY PJ (**1997**). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23 (1-3), 3–25.

[50] CLARK DE & PICKETT SD (**2000**). Computational methods for the prediction of 'drug-likeness'. *Drug Discovery Today*, 5 (2), 49–58.

[51] CONGREVE M, CARR R, MURRAY C, & JHOTI H (**2003**). A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today*, 8 (19), 876–877.

[52] MALY DJ, CHOONG IC, & ELLMAN JA (**2000**). Combinatorial target-guided ligand assembly: Identification of potent subtype-selective c-Src inhibitors. *PNAS*, 97 (6), 2419–2424.

[53] DALVIT C, ARDINI E, FLOCCO M, FOGLIATTO G, MONGELLI N, & VERONESI M (**2003**). A General NMR Method for Rapid, Efficient, and Reliable Biochemical Screening. *Journal of the American Chemical Society*, 125 (47), 14620–14625.

[54] DALVIT C, ARDINI E, FOGLIATTO G, MONGELLI N, & VERONESI M (**2004**). Reliable high-throughput functional screening with 3-FABS. *Drug Discovery Today*, 9 (14), 595–602.

[55] HARTSHORN M, MURRAY C, CLEASBY A, FREDERICKSON M, TICKLE I, & JHOTI H (**2005**). Fragment-based lead discovery using X-ray crystallography. *Journal of Medicinal Chemistry*, 48 (2), 403–413.

[56] SHARFF A & JHOTI H (**2003**). High-throughput crystallography to enhance drug discovery. *Current Opinion in Chemical Biology*, 7, 340–345.

[57] MATTOS C & RINGE D (**1996**). Locating and characterizing binding sites on proteins. *Nature Biotechnology*, 14 (5), 595–599.

[58] MOUNTAIN V (**2003**). Astex, Structural Genomix, and Syrrx. I can see clearly now: Structural Biology and Drug Discovery. *Chemistry & Biology*, 10, 95–98.

[59] NIENABER VL, RICHARDSON PL, KLIGHOFER V, BOUSKA JJ, GIRANDA VL, & GREER J (**2000**). Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nature Biotechnology*, 28, 1105–1106.

[60] TRIGGLE D (**2003**). Medicines in the 21st Century or Pills, Politics, Potions, and Profits: Where is Public Policy. *Drug Development Research*, 59, 269–291.

[61] LESUISSE D, LANGE G, DEPREZ P, BENARD D, SCHOOT B, DELETTRE G, MARQUETTE JP, BROTO P, JEAN-BAPTISTE V, BICHET P, SARUBBI E, & MANDINE E (**2002**). SAR and X-ray. A new approach combining fragment-based screening and rational drug design: Application to the discovery of nanomolar inhibitors of Src SH2. *Journal of Medicinal Chemistry*, 45 (12), 2379–2387.

[62] SWAYZE E, JEFFERSON E, SANNES-LOWERY K, BLYN L, RISEN L, S A, OSGOOD S, HOFSTADLER S, & GRIFFEY R (**2002**). SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *Journal of Medicinal Chemistry*, 45 (18), 3816–3819.

[63] OCKEY D, DOTSON J, STRUBLE M, STULTS J, BOURELL J, CLARK K, & GADEK T (**2004**). Structure-activity relationships by mass spectrometry: identification of novel MMP-3 inhibitors. *Bioorganic and Medicinal Chemistry*, 12 (1), 37–44.

[64] BREUKER K (**2004**). New mass spectrometric methods for the quantification of protein-ligand binding in solution. *Angewandte Chemie International Edition*, 43, 22–25.

[65] ERLANSON DA, WELLS J, & BRAISTED A (**2004**). Tethering: Fragment-based drug discovery. *Annual Review of Biophysics and Biomolecular structure*, 33, 199–223.

[66] HAJDUK PJ, MEADOWS RP, & FESIK SW (**1999**). NMR-based screening in drug discovery. *Quarterly Reviews of Biophysics*, 32 (3), 211–240.

[67] COLES M, HELLER M, & KESSLER H (**2003**). NMR-based screening technologies. *Drug Discovery Today*, 8 (17), 803–810.

[68] MEYER B & PETERS T (**2003**). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition*, 42 (8), 865–890.

[69] LEPRE CA, MOORE JM, & PENG JW (**2004**). Theory and applications of NMR-based screening in pharmaceutical research. *Chemical Reviews*, 104 (8), 3641–3676.

[70] CORNILESCU G, DELAGLIO F, & BAX A (**1999**). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, 13, 289–302.

[71] NEAL S, NIP A, ZHANG H, & WISHART D (**2003**). Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *Journal of Biomolecular NMR*, 26, 215–240.

[72] OVERHAUSER AW (**1953**). Paramagnetic Relaxation in Metals. *Physical Review*, 89, 689–700.

[73] MEIBOOM S & GILL D (**1958**). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Reviews of Scientific Instruments*, 29 (8), 688–691.

[74] HAJDUK PJ, OLEJNICZAK ET, & FESIK SW (**1997**). One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *Journal of the American Chemical Society*, 119 (50), 12257–12261.

[75] JAHNKE W, PEREZ LB, PARIS CG, STRAUSS A, FENDRICH G, & NALIN CM (**2000**). Second-site NMR screening with a spin-labeled first ligand. *Journal of the American Chemical Society*, 122 (30), 7394–7395.

[76] JAHNKE W, RUDISSER S, & ZURINI M (**2001**). Spin label enhanced NMR screening. *Journal of the American Chemical Society*, 123 (13), 3149–3150.

[77] VANWETSWINKEL S, HEETEBRIJ R, VAN DUYNHOVEN J, HOLLANDER J, FILIPPOV D, HAJDUK P, & SIEGAL G (**2005**). TINS, target immobilized NMR screening: an efficient and sensitive method for ligand discovery. *Chemistry & Biology*, 12 (2), 207–216.

[78] LIN M, SHAPIRO M, & WAREING J (**1997**). Diffusion-edited NMR - Affinity NMR for direct observation of molecular interactions. *Journal of the American Chemical Society*, 119, 5249–5250.

[79] STEJSKAL EO & TANNER JE (**1965**). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *Journal of Chemical Physics*, 42 (1), 288–292.

[80] CHEN A & SHAPIRO M (**1999**). A new drug-screening tool that probes ligand-receptor interactions. *Analytical Chemistry News & Features*, pp. 669A–675A.

[81] MAYER M & MEYER B (**1999**). Characterisation of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angewandte Chemie International Edition*, 38 (12), 1784–1787.

[82] KLEIN J, MEINECKE R, MAYER M, & MEYER B (**1999**). Detecting Binding Affinity to Immobilized Receptor Proteins in Compound Libraries by HR-MAS STD NMR. *Journal of the American Chemical Society*, 121 (22), 5336–5337.

[83] MEINECKE R & MEYER B (**2001**). Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin αIIbβ3. *Journal of Medicinal Chemistry*, 44 (19), 3059–3065.

[84] BENIE A, MOSER R, BAUML E, BLAAS D, & PETERS T (**2003**). Virus-ligand interactions: identification and characterization of ligand binding by NMR spectroscopy. *Journal of the American Chemical Society*, 125 (1), 14–15.

[85] BIET T & PETERS T (**2001**). Molecular Recognition of UDP-Gal by β-1,4-Galactosyltransferase T1. *Angewandte Chemie International Edition*, 40 (22), 4189–4192.

[86] JAYALAKSHMI V, BIET T, PETERS T, & KRISHNA NR (**2004**). Refinement of the Conformation of UDP-Galactose Bound to Galactosyltransferase Using the STD NMR Intensity-Restrained CORCEMA Optimization. *Journal of the American Chemical Society*, 126 (28), 8610–8611.

[87] DALVIT C, PEVARELLO P, TATÒ M, VERONESI M, VULPETTI A, & SUNDSTRÖM M (**2000**). Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *Journal of Biomolecular NMR*, 18, 65–68.

[88] DALVIT C, FOGLIATTO G, STEWART A, VERONESI M, & STOCKMAN B (**2001**). WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular NMR*, 21 (4), 349–359.

[89] CHEN A & SHAPIRO M (**2000**). NOE Pumping. 2. A High-Throughput Method to Determine Compounds with Binding Affinity to Macromolecules by NMR. *Journal of the American Chemical Society*, 122, 414–415.

[90] DALVIT C, FLOCCO M, KNAPP S, MOSTARDINI M, PEREGO R, STOCKMAN B, VERONESI M, & VARASI M (**2002**). High-Throughput NMR-Based Screening with Competition Binding Experiment. *Journal of the American Chemical Society*, 124, 7702–7709.

[91] SIRIWARDENA A, TIAN F, NOBLE S, & PRESTEGARD J (**2002**). A straightforward NMR-spectroscopy-based method for rapid library screening. *Angewandte Chemie International Edition*, 41 (18), 3454–3457.

[92] JAHNKE W, FLOERSHEIM P, OSTERMEIER C, ZHANG X, HEMMIG R, HURTH K, & UZUNOV D (**2002**). NMR reporter screening for the detection of high-affinity ligands. *Angewandte Chemie International Edition*, 41 (18), 3420–3423.

[93] DALVIT C, FAGERNESS PE, HADDEN DTA, SARVER RW, & STOCKMAN BJ (**2003**). Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *Journal of the American Chemical Society*, 125, 7696–7703.

[94] WANG YS, LIU D, & WYSS D (**2004**). Competition STD NMR for the detection of high-affinity ligand and NMR-based screening. *Magnetic Resonance in Chemistry*, 42, 485–489.

[95] DALVIT C, FASOLINI M, FLOCCO M, KNAPP S, PEVARELLO P, & VERONESI M (**2002**). NMR-based screening with competition water-ligand observed via gradient spectroscopy experiments: detection of high-affinity ligands. *Journal of Medicinal Chemistry*, 45, 2610–2614.

[96] PERVUSHIN K, RIEK R, WIDER G, & WÜTHRICH K (**1997**). Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences U.S.A.*, 94 (23), 12366–12371.

[97] HAJDUK PJ, GERFIN T, BOEHLEN JM, HABERLI M, MAREK D, & FESIK SW (**1999**). High-throughput nuclear magnetic resonance-based screening. *Journal of Medicinal Chemistry*, 42 (13), 2315–2317.

[98] ZARTLER E, HANSON J, JONES B, KLINE A, MARTIN G, MO H, SHAPIRO M, WANG R, WU H, & YAN J (**2003**). RAMPED-UP NMR: Multiplexed NMR-based screening for drug discovery. *Journal of the American Chemical Society*, 125, 10941–10946.

[99] CAMPBELL SJ & JACKSON RM (**2003**). Diversity in the SH2 domain family phosphotyrosyl peptide binding site. *Protein Engineering*, 16 (3), 217–227.

[100] WAKSMAN G, KOMINOS D, ROBERTSON SC, PANT N, BALTIMORE D, BIRGE RB, COWBURN D, HANAFUSA H, MAYER BJ, OVERDUIN M, RESH MD, RIOS CB, SILVERMAN L, & KURIYAN J (**1992**). Crystal-Structure of the Phosphotyrosine Recognition Domain SH2 of V-Src Complexed with Tyrosine-Phosphorylated Peptides. *Nature*, 358 (6388), 646–653.

[101] WAKSMAN G, SHOELSON SE, PANT N, COWBURN D, & KURIYAN J (**1993**). Binding of a High-Affinity Phosphotyrosyl Peptide to the Src SH2 Domain - Crystal-Structures of the Complexed and Peptide-Free Forms. *Cell*, 72 (5), 779–790.

[102] ETTMAYER P, FRANCE D, GOUNARIDES J, JAROSINSKI M, MARTIN MS, RONDEAU JM, SABIO M, TOPIOL S, WEIDMANN B, ZURINI M, & BAIR KW (**1999**). Structural and Conformational Requirements for High-Affinity Binding to the SH2 Domain of Grb2. *Journal of Medicinal Chemistry*, 42 (6), 971–980.

[103] GALLIVAN J & DOUGHERTY D (**1999**). Cation-π Interactions in Structural Biology. *Proceedings of the National Academy of Scientists, USA*, 96, 9459–9464.

[104] LEMMON MA & LADBURY JE (**1994**). Thermodynamic Studies of Tyrosyl-Phosphopeptide Binding to the SH2 Domain of P56(Lck). *Biochemistry*, 33 (17), 5070–5076.

[105] BRADSHAW JM, GRUCZA RA, LADBURY JE, & WAKSMAN G (**1998**). Probing the"two-pronged plug two-holed socket" model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: A thermodynamic study. *Biochemistry*, 37 (25), 9083–9090.

[106] BRADSHAW JM, MITAXOV V, & WAKSMAN G (**1999**). Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. *Journal of Molecular Biology*, 293 (4), 971–985.

[107] BRADSHAW JM, MITAXOV V, & WAKSMAN G (**2000**). Mutational investigation of the specificity determining region of the Src SH2 domain. *Journal of Molecular Biology*, 299 (2), 521–535.

[108] LUBMAN OY & WAKSMAN G (**2002**). Dissection of the energetic coupling across the Src SH2 domain-tyrosyl phosphopeptide interface. *Journal of Molecular Biology*, 316 (2), 291–304.

[109] PICCIONE E, CASE R, DOMCHEK S, HU P, CHAUDHURI M, BACKER J, SCHLESSINGER J, & SHOELSON S (**1993**). Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding. *Biochemistry*, 32 (13), 3197–3202.

[110] SONGYANG Z, SHOELSON SE, CHAUDHURI M, GISH G, PAWSON T, HASER WG, KING F, ROBERTS T, RATNOFSKY S, LECHLEIDER RJ, NEEL B, BIRGE R, FAJARDO J, CHOU M, HANAFUSA H, SCHAFFHAUSEN B, & CANTLEY L (**1993**). SH2 domains recognize specific phosphopeptide sequences. *Cell*, 72 (5), 767–78.

[111] SONGYANG Z, SHOELSON S, MCGLADE J, OLIVIER P, PAWSON T, BUSTELO X, BARBACID M, SABE H, HANAFUSA H, & YI T (**1994**). Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Molecular and Cellular Biology*, 14 (4), 2777–2785.

[112] CANTLEY L & SONGYANG Z (**1994**). Specificity in recognition of phosphopeptides by src-homology 2 domains. *Journal of Cell Science Supplement*, 18, 121–126.

[113] YAFFE M, LEPARC G, LAI J, OBATA T, VOLINA S, & CANTLEY L (**2001**). A motif-based profile scanning approach for genome-wide prediction of signalling pathways. *Nature Biotechnology*, 19, 348–353.

[114] MARENGERE LEM, ZHOU SY, GISH GD, SCHALLER MD, PARSONS JT, STERN MJ, CANTLEY LC, & PAWSON T (**1994**). SH2 Domain Specificity and Activity Modified by a Single Residue. *Nature*, 369 (6480), 502–505.

[115] HUYER G & RAMACHANDRAN C (**1998**). The specificity of the N-terminal SH2 domain of SHP-2 is modified by a single point mutation. *Biochemistry*, 37 (9), 2741–2747.

[116] LADBURY J (**2005**). "Protein-Protein Recognition in Phosphotyrosine-Mediated Intracellular Signalling", in Proteomics and Protein-Protein Interactions: Biology, Chemistry, Bioinformatics, And Drug Design. In press.

[117] HENRIQUES DA & LADBURY JE (**2001**). Inhibitors to the Src SH2 domain: A lesson in Structure-Thermodynamic Correlation in Drug Design. *Archives of Biochemistry and Biophysics*, 390 (2), 158–168.

[118] O'BRIEN R, RUGMAN P, RENZONI D, LAYTON M, HANDA R, HILYARD K, WATERFIELD MD, DRISCOLL PC, & LADBURY JE (**2000**). Alternative modes of binding of proteins with tandem SH2 domains. *Protein Science*, 9 (3), 570–579.

[119] O'ROURKE L & LADBURY JE (**2003**). Specificity is complex and time consuming: Mutual exclusivity in tyrosine kinase-mediated signaling. *Accounts of Chemical Research*, 36 (6), 410–416.

[120] PAWSON T (**2004**). Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems. *Cell*, 116, 191–203.

[121] HUBBARD SR (**1999**). Src autoinhibition: let us count the ways. *Nature Structural Biology*, 6 (8), 711–714.

[122] COOPER J, GOULD K, CARTWRIGHT C, & HUNTER T (**1986**). Tyr527 is phosphorylated in pp60c-src: implications for regulation. *Science*, 231, 1431–1434.

[123] NADA S, OKADA M, MACAULEY A, COOPER J, & NAKAGAWA H (**1991**). Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60c-src. *Nature*, 351 (6321), 69–72.

[124] HARDER K, MOLLER N, PEACOCK J, & JIRIK F (**1998**). Protein-tyrosine phosphatase alpha regulates Src family kinases and alters cell-substratum adhesion. *Journal of Biological Chemistry*, 273, 31890–31900.

[125] SU J, MURANJAN M, & SAP J (**1999**). Receptor protein tyrosine phosphatase alpha activates Src-family kinases and controls integrin-mediated responses in fibroblasts. *Current Biology*, 9, 505–511.

[126] PONNNIAH S, WANG D, LIM K, & PALLEN C (**1999**). Targeted disruption of the tyrosine phosphatase PTPalpha leads to constitutive downregulation of the kinases Src and Fyn. *Current Biology*, 9, 535–538.

[127] XU W, HARRISON S, & ECK M (**1997**). Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, 385 (6617), 595–602.

[128] XU W, DOSHI A, LEI M, ECK M, & HARRISON S (**1999**). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Molecular Cell*, 3, 629–638.

[129] SICHERI F, MOAREFI I, & KURIYAN J (**1997**). Crystal structure of the Src family tyrosine kinase Hck. *Nature*, 385, 602–609.

[130] GONFLONI S, FRISCHKNECHT F, WAY M, & SUPERTI-FURGA G (**1999**). Leucine 255 of Src couples intramolecular interactions to inhibition of catalysis. *Nature Structural Biology*, 6 (8), 760–764.

[131] LUBMAN OY & WAKSMAN G (**2003**). Structural and thermodynamic basis for the interaction of the Src SH2 domain with the activated form of the PDGF beta-receptor. *Journal of Molecular Biology*, 328 (3), 655–668.

[132] GONFLONI S, WEIJLAND A, KRETZSCHMAR J, & SUPERTI-FURGA G (**2000**). Crosstalk between the catalytic and regulatory domains allows bidirectional regulation of Src. *Nature Structural Biology*, 7 (4), 281–286.

[133] HARDWICK S JAMES & SEFTON BM (**1997**). The Activated Form of the Lck Tyrosine Protein Kinase in Cells Exposed to Hydrogen Peroxide Is Phosphorylated at Both Tyr-394 and Tyr-505. *Journal of Biological Chemistry*, 272, 25429–25432.

[134] SUN G, SHARMA A, & BUDDE R (**1998**). Autophosphorylation of Src and Yes blocks their inactivation by Csk phosphorylation. *Oncogene*, 17 (12), 1587–1595.

[135] SOTIRELLIS N, JOHNSON TM, HIBBS ML, STANLEY IJ, STANLEY E, DUNN AR, & CHENG HC (**1995**). Autophosphorylation Induces Autoactivation and a Decrease in the Src Homology 2 Domain Accessibility of the Lyn Protein Kinase. *Journal of Biological Chemistry*, 270 (50), 29773–29780.

[136] FUKAMI Y, SATO K, IKEDA K, KAMISANGO K, KOIZUMI K, & MATSUNO T (**1993**). Evidence for autoinhibitory regulation of the c-src gene product. A possible interaction between the src homology 2 domain and autophosphorylation site. *Journal of Biological Chemistry*, 268 (2), 1132–1140.

[137] BROWN M & COOPER J (**1996**). Regulation, substrates and functions of src. *Biochimica et Biophysica Acta*, 1287, 121–149.

[138] ABRAM C & COURTNEIDGE S (**2000**). Src family tyrosine kinases and growth factor signalling. *Experimental Cell Research*, 254, 1–13.

[139] ROUS P (**1911**). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *Journal of Experimental Medicine*, 13, 397–411.

[140] MARTIN G (**2004**). The road to Src. *Oncogene*, 23 (48), 7910–7917.

[141] HANAFUSA H, HANAFUSA T, & RUBIN H (**1963**). The defectiveness of Rous sarcoma virus. *Proceedings of the National Academy of Sciences*, 49, 572–580.

[142] BRUGGE J & ERIKSON R (**1977**). Identification of a transformation-specific antigen induced by an avian sarcoma virus. *Nature*, 269, 346–348.

[143] PURCHIO A, ERIKSON E, BRUGGE J, & ERIKSON R (**1978**). Identification of a polypeptide encoded by the avian sarcoma virus src gene. *Proceedings of the National Academy of Sciences U.S.A.*, 75 (3), 1567–1571.

[144] COLLETT M & ERIKSON R (**1978**). Protein kinase activity associated with the avian sarcoma virus src gene product. *Proceedings of the National Academy of Sciences U.S.A.*, 75 (4), 2021–2024.

[145] FRAME M (**2002**). Src in cancer: deregulation and consequences for cell behaviour. *Biochimica et Biophysica Acta*, 1602, 114–130.

[146] ALPER O & BOWDEN E (**2005**). Novel insights into c-Src. *Current Pharmaceutical Design*, 11 (9), 1119–1130.

[147] IRBY R, MAO W, COPPOLA D, KANG J, LOUBEAU J, TRUDEAU W, KARL R, FUJITA D, JOVE R, & YEATMAN T (**1999**). Activating SRC mutation in a subset of advanced human colon cancers. *Nature Genetics*, 21 (2), 187–190.

[148] IMAMOTO A & SORIANO P (**1993**). Disruption of the csk gene, encoding a negative regulator of Src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice. *Cell*, 73, 1117–1124.

[149] NADA S, YAGI T, TAKEDA H, TOKUNAGA T, NAKAGAWA H, IKAWA Y, OKADA M, & AIZAWA S (**1993**). Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell*, 73, 1125–1135.

[150] BAIROCH A, BOECKMANN B, FERRO S, & GASTEIGER E (**2004**). Swiss-Prot: Juggling between evolution and stability. *Briefings in Bioinformatics*, 5, 39–55.

[151] SORIANO P, MONTGOMERY C, GESKE R, & ALLAN B (**1991**). Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell*, 64 (4), 693–702.

[152] HORNE W, NEFF L, CHATTERJEE D, LOMRI A, LEVY J, & BARON R (**1992**). Osteoclasts express high levels of pp60c-src in association with intracellular membranes. *Journal of Cell Biology*, 119 (4), 1003–1014.

[153] LOWELL C, NIWA M, SORIANO P, & VARMUS H (**1996**). Deficiency of the Hck and Src tyrosine kinases results in extreme levels of extramedullary hematopoiesis. *Blood*, 87 (5), 1780.

[154] SCHWARTZBERG PL, XING L, HOFFMAN O, LOWELL CA, GARRETT L, BOYCE BF, & VARMUS HE (**1997**). Rescue of osteoclast function by transgenic expression of kinase-deficient Src in *src-/-* mutant mice. *Genes Development*, 11, 2835–2844.

[155] BOYCE B, YONEDA T, LOWE C, SORIANO P, & MUNDY G (**1992**). Requirement of pp60c-src expression for osteoclasts to form ruffled borders and resorb bone in mice. *Journal of Clinical Investigation*, 90, 1622–1627.

[156] LOWE C, YONEDA T, BOYCE B, CHEN H, MUNDY G, & SORIANO P (**1993**). Osteopetrosis in *src*-deficient mice is due to an autonomous defect of osteoclasts. *Proceedings of the National Academy of Sciences U.S.A.*, 90, 4485–4489.

[157] WATTS N (**2003**). Bisphosphonate treatment of osteoporosis. *Clinical Geriatric Medicine*, 19, 395–414.

[158] LICATA A (**2005**). Discovery, clinical development, and therapeutic uses of bisphosphonates. *Annals Of Pharmacotherapy*, 39 (4), 668–677.

[159] SHAKESPEARE W, YANG M, BOHACEK R, CERASOLI F, STEBBINS K, SUNDARAMOORTHI R, AZIMIOARA M, VU C, PRADEEPAN S, METCALF C III, HARALDSON C, MERRY T, DALGARNO D, NARULA S, HATADA M, LU X, VAN SCHRAVENDIJK MR, ADAMS S, VIOLETTE S, SMITH J, GUAN W, BARTLETT C, HERSON J, IULIUCCI J, WEIGELE M, & SAWYER T (**2000**). Structure-based design of an osteoclast-selective, nonpeptide Src homology 2 inhibitor with *in vivo* antiresorptive activity. *PNAS*, 97 (17), 9373–9378.

[160] VIOLETTE SM, SHAKESPEARE WC, BARTLETT C, GUAN W, SMITH JA, RICKLES RJ, BOHACEK RS, HOLT DA, BARON R, & SAWYER TK (**2000**). A Src SH2 selective binding compound inhibits osteoclast-mediated resorption. *Chemistry & Biology*, 7 (3), 225–235.

[161] VIOLETTE SM, GUAN W, BARTLETT C, SMITH JA, BARDELAY C, ANTOINE E, RICKLES RJ, MANDINE E, VAN SCHRAVENDIJK MR, ADAMS SE, LYNCH BA, SHAKESPEARE WC, YANG M, JACOBSEN VA, TAKEUCHI CS, MACEK KJ, BOHACEK RS, DALGARNO DC, WEIGELE M, LESUISSE D, SAWYER TK, & BARON R (**2001**). Bone-targeted Src SH2 inhibitors block Src cellular activity and osteoclast-mediated resorption. *Bone*, 28 (1), 54–64.

[162] STRAUS D & WEISS A (**1992**). Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell*, 70, 585–593.

[163] MOLINA T, KISHIHARA K, SIDEROVSKI D, VAN EWIJK W, NARENDRAN A, TIMMS E, WAKEHAM A, PAIGE C, HARTMANN K, VEILLETTE A, DAVIDSON D, & MAK T (**1992**). Profound block in thymocyte development in mice lacking p56lck. *Nature*, 357, 161–164.

[164] SAWYER T, BOYCE B, DALGARNO D, & IULIUCCI J (**2001**). Src inhibitors: genomics to therapeutics. *Expert Opin. Investig. Drugs*, 10 (7), 1327–1344.

[165] LIU X, BRODEUR S, GISH G, SONGYANG Z, CANTLEY L, LAUDANO A, & PAWSON T (**1993**). Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. *Oncogene*, 8, 1119–1126.

[166] ALONSO G, KOEGL M, MAZURENKO M, & COURTNEIDGE S (**1995**). Sequence requirements for binding of Src family tyrosine kinases to activated growth factors. *Journal of Biological Chemistry*, 270, 9840–9848.

[167] BOERNER R, KASSEL D, BARKER S, ELLIS B, DELACY P, & KNIGHT W (**1996**). Correlation of the phosphorylation states of pp60c-src with tyrosine kinase activity: the intramolecular pY530-SH2 complex retains significant activity if Y419 is phosphorylated. *Biochemistry*, 35, 9519–9525.

[168] MANDINE E, JEAN-BAPTISTE V, VAYSSIERE W, GOFFLO D, BENARD D, SARUBBI E, DEPREZ P, BARON R, SUPERTI-FURGA G, & LESUISSE D (**2002**). High-affinity Src-SH2 ligands which do not activate Tyr(527)-phosphorylated Src in an experimental in vivo system. *Biochemical and Biophysical Research Communications*, 298 (2), 185–192.

[169] SONATORE L, WISNIEWSKI D, FRANK L, CAMERON P, HERMES J, MARCY A, & SALOWE S (**1996**). The utility of FK506-binding protein as a fusion partner in scintillation proximity assays: application to SH2 domains. *Analytical Biochemistry*, 240 (2), 289–297.

[170] GAY B, SUAREZ S, CARAVATTI G, FURET P, MEYER T, & SCHOEPFER J (**1999**). Selective Grb2 SH2 inhibitors as anti-ras therapy. *International Journal of Cancer*, 83, 235–241.

[171] LYNCH B, LOIACONO K, TIONG C, ADAMS S, & MACNEIL I (**1997**). A fluorescence polarization based Src-SH2 binding assay. *Analytical Biochemistry*, 247 (1), 77–82.

[172] MANDINE E, GOFFLO D, JEAN-BAPTISTE V, SARUBBI E, TOUYER G, DEPREZ P, & LESUISSE D (**2001**). Src homology-2 domain binding assays by scintillation proximity and surface plasmon resonance. *Journal of Molecular Recognition*, 14 (4), 254–260.

[173] HAJDUK PJ, ZHOU MM, & FESIK SW (**1999**). NMR-based discovery of phosphotyrosine mimetics that bind to the Lck SH2 domain. *Bioorganic & Medicinal Chemistry Letters*, 9 (16), 2403–2406.

[174] RICKLES RJ, HENRY PA, GUAN W, AZIMIOARA M, SHAKESPEARE WC, VIOLETTE S, & ZOLLER MJ (**1998**). A novel mechanism-based mammalian cell assay for the identification of SH2-domain specific protein-protein inhibitors. *Chemistry & Biology*, 5 (10), 529–538.

[175] GILMER T, RODRIQUEZ M, JORDAN S, CROSBY R, ALLIGOOD K, GREEN M, KIMERY M, WAGNER C, KINDER D, CHARIFSON P, HASSELL AM, WILLARD D, LUTHER M, RUSNAK D, STERNBACH DD, MEHROTRA M, PEEL M, SHAMPINE L, DAVIS R, ROBBINS J, PATEL IR, KASSEL D, BURKHART W, MOYER M, BRADSHAW T, & BERMAN J (**1994**). Peptide Inhibitors of Src SH3-SH2-Phosphoprotein Interactions. *Journal of Biological Chemistry*, 269 (50), 31711–31719.

[176] DOMCHEK S, AUGER K, CHATTERJEE S, BURKE J TR, & SHOELSON S (**1992**). Inhibition of SH2 domain/phosphoprotein association by nonhydrolyzable phosphonopeptide. *Biochemistry*, 31, 9865–9870.

[177] STANKOVIC CJ, SURENDRAN N, LUNNEY EA, PLUMMER MS, PARA KS, SHAHRIPOUR A, FERGUS JH, MARKS JS, HERRERA R, HUBBELL SE, HUMBLET C, SALTIEL AR, STEWART BH, & SAWYER TK (**1997**). The role of 4-phosphonodifluoromethyl- and 4-phosphonophenylalanine in the selectivity and cellular uptake of SH2 domain ligands. *Bioorganic & Medicinal Chemistry Letters*, 7 (14), 1909–1914.

[178] BURKE T JR, SMYTH S, OTAKA A, NOMIZU M, ROLLER P, WOLF G, CASE R, & SHOELSON S (**1994**). Nonhydrolyzable Phosphotyrosyl Mimetics for the Preparation Of Phosphatase-Resistant SH2 Domain Inhibitors. *Biochemistry*, 33, 6490–6494.

[179] EATON SR, CODY WL, DOHERTY AM, HOLLAND DR, PANEK RL, LU GH, DAHRING TK, & ROSE DR (**1998**). Design of Peptidomimetics That Inhibit the Association of Phosphatidylinositol 3-Kinase with Platelet-Derived Growth Factor-β Receptor and Possess Cellular Activity. *Journal of Medicinal Chemistry*, 41 (22), 4329–4342.

[180] DEPREZ P, MANDINE E, GOFFO D, MEUNIER S, & LESUISSE D (**2002**). Small ligands interacting with the phosphotyrosine binding pocket of the Src SH2 protein. *Bioorganic and Medicinal Chemistry Letters*, 12, 1295–1298.

[181] LANGE G, LESUISSE D, DEPREZ P, SCHOOT B, LOENZE P, BÉNARD D, MARQUETTE JP, BROTO P, SARUBBI E, & MANDINE E (**2003**). Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of pp60Src are identical to those for high affinity binding of full length inhibitors. *Journal of Medicinal Chemistry*, 46, 5184–5195.

[182] LIU WQ, ROQUES B, & GARBAY C (**1997**). Synthesis of L-2,3,5,6-tetrafluoro-4-(phosphonomethyl)phenylalanine, a novel non-hydrolyzable phosphotyrosine mimetic and L-4-(phosphonodifluoro-methyl)phenylalanine. *Tetrahedron Letters*, 38 (8), 1389–1392.

[183] FU JM & CASTELHANO AL (**1998**). Design and synthesis of a pyridone-based phosphotyrosine mimetic. *Bioorganic and Medicinal Chemistry Letters*, 8 (19), 2813–2816.

[184] BEAULIEU PL, CAMERON DR, FERLAND JM, GAUTHIER J, GHIRO E, GILLARD J, GORYS V, POIRIER M, RANCOURT J, WERNIC D, LLINAS-BRUNET M, BETAGERI R, CARDOZO M, HICKEY ER, INGRAHAM R, JAKES S, KABCENELL A, KIRRANE T, LUKAS S, PATEL U, PROUDFOOT J, SHARMA R, TONG L, & MOSS N (**1999**). Ligands for the Tyrosine Kinase p56lck SH2 Domain: Discovery of Potent Dipeptide Derivatives with Monocharged, Nonhydrolyzable Phosphate Replacements. *Journal of Medicinal Chemistry*, 42 (10), 1757–1766.

[185] MEHROTRA MM, STERNBACH DD, RODRIGUEZ M, CHARIFSON P, & BERMAN J (**1996**). α-Dicarbonyls as "non-charged" arginine-directed affinity labels. Novel synthetic routes to α-dicarbonyl analogs of the pp60c-src SH2 domain-targeted phosphopeptide Ac-Tyr(OPO3H2)-Glu-Glu-Ile-Glu. *Bioorganic and Medicinal Chemistry Letters*, 9, 1941–1946.

[186] MILLER M, BRACCOLINO D, CLEARY D, REAM J, WALKER M, & SIKORSKI J (**1994**). EPSP synthase inhibitor design IV. New aromatic substrate analogues and symmetrical inhibitors containing a novel 3-phosphate mimic. *Bioorganic and Medicinal Chemistry Letters*, 4, 2605–2608.

[187] TONG L, WARREN TC, LUKAS S, SCHEMBRI-KING J, BETAGERI R, PROUDFOOT JR, & JAKES S (**1998**). Carboxymethyl-phenylalanine as a Replacement for Phosphotyrosine in SH2 Domain Binding. *Journal of Biological Chemistry*, 273, 20238–20242.

[188] BURKE T JR, LUO J, YAO ZJ, GAO Y, ZHAO H, MILNE GWA, GUO R, VOIGT JH, KING CR, & YANG D (**1999**). Monocarboxylic-based phosphotyrosyl mimetics in the design of GRB2 SH2 domain inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 9, 347–352.

[189] YAO ZJ, GAO Y, VOIGT JH, FORD H JR, & BURKE TR JR (**1999**). Synthesis of Fmoc-protected 4-carboxydifluoromethyl-L-phenylalanine: A phosphotyrosyl mimetic of potential use for signal transduction studies. *Tetrahedron*, 55 (10), 2865–2874.

[190] BURKE T JR, YE B, AKAMATSU M, FORD H JR, YAN X, KOLE H, WOLF G, SHOELSON S, & ROLLER P (**1999**). 4'-O-[2-(2-Fluoromalonyl)]-L-tyrosine: A Phosphotyrosyl Mimic for the Preparation of Signal Transduction Inhibitory Peptides. *Journal of Medicinal Chemistry*, 39 (5), 1021–1027.

[191] GAO Y, LUO J, YAO ZJ, GUO R, ZOU H, KELLEY J, VOIGT J, YANG D, & BURKE T JR (**2000**). Inhibition of Grb2 SH2 Domain Binding by Non-Phosphate-Containing Ligands. 2. 4-(2-Malonyl)phenylalanine as a Potent Phosphotyrosyl Mimetic. *Journal of Medicinal Chemistry*, 43 (4), 911–920.

[192] SUNDARAMOORTHI R, KAWAHATA N, YANG M, SHAKESPEARE W, METCALF R CA, WANG Y, MERRY T, EYERMANN C, BOHACEK R, NARULA S, DALGARNO D, & SAWYER T (**2003**). Structure-based design of novel nonpeptide inhibitors of the Src SH2 domain:phosphotyrosine mimetics exploiting multifunctional group replacement chemistry. *Biopolymers*, 71, 717–729.

[193] BRADSHAW JM & WAKSMAN G (**1998**). Calorimetric investigation of proton linkage by monitoring both the enthalpy and association constant of binding: Application to the interaction of the Src SH2 domain with a high-affinity tyrosyl phosphopeptide. *Biochemistry*, 37 (44), 15400–15407.

[194] CHARIFSON PS, SHEWCHUK LM, ROCQUE W, HUMMEL CW, JORDAN SR, MOHR C, PACOFSKY GJ, PEEL MR, RODRIGUEZ M, STERNBACH DD, & CONSLER TG (**1997**). Peptide ligands of pp60$^{(c-src)}$ SH2 domains: A thermodynamic and structural study. *Biochemistry*, 36 (21), 6283–6293.

[195] ALLIGOOD KJ, CHARIFSON PS, CROSBY R, CONSLER TG, FELDMAN PL, GAMPE RT, GILMER TM, JORDAN SR, MILSTEAD MW, MOHR C, PEEL MR, ROCQUE W, RODRIGUEZ M, RUSNAK DW, SHEWCHUK LM, & STERNBACH DD (**1998**). The formation of a covalent complex between a dipeptide ligand and the src SH2 domain. *Bioorganic & Medicinal Chemistry Letters*, 8 (10), 1189–1194.

[196] KAWAHATA N, YANG MG, LUKE GP, SHAKESPEARE WC, SUNDARAMOORTHI R, WANG YH, JOHNSON D, MERRY T, VIOLETTE S, GUAN W, BARTLETT C, SMITH J, HATADA M, LU XD, DALGARNO DC, EYERMANN CJ, BOHACEK RS, & SAWYER TK (**2001**). A novel phosphotyrosine mimetic 4'-carboxymethyloxy-3'-phosphonophenylalanine (Cpp): Exploitation in the design of nonpeptide inhibitors of pp60(Src) SH2 domain. *Bioorganic & Medicinal Chemistry Letters*, 11 (17), 2319–2323.

[197] SAWYER TK (**1998**). Src homology-2 Domains: Structure, Mechanisms, and Drug Discovery. *Biopolymers*, 47, 243–261.

[198] NAM N, PITTS R, SUN G, SARDARI S, TIEMO A, XIE M, YAN B, & PARANG K (**2004**). Design of tetrapeptide ligands as inhibitors of the Src SH2 domain. *Bioorganic and Medicinal Chemistry Letters*, 12, 779–787.

[199] (PATENT) (**1998**). New inhibitors of SH2-mediated processes. *Expert Opinion on Therapeutic Patents*, 8 (3), 333–335.

[200] LUNNEY EA, PARA KS, RUBIN JR, HUMBLET C, FERGUS JH, MARKS JS, & SAWYER TK (**1997**). Structure-based design of a novel series of nonpeptide ligands that bind to the pp60(src) SH2 domain. *Journal of the American Chemical Society*, 119 (51), 12471–12476.

[201] YEH R, LEE T, & LAWRENCE D (**2004**). From consensus sequence peptide to high affinity ligand, a "library-scan" strategy. *Journal of Biological Chemistry*, 42, 784–787.

[202] PARK SH, WON J, & LEE KH (**2002**). Design and characterization of non-phosphopeptide inhibitors for Src family SH2 domains. *Bioorganic & Medicinal Chemistry Letters*, 12 (19), 2711–2714.

[203] LEE TR & LAWRENCE DS (**1999**). Acquisition of high-affinity, SH2-targeted ligands via a spatially focused library. *Journal of Medicinal Chemistry*, 42 (5), 784–787.

[204] BALL J & ALEWOOD P (**1990**). Conformational constraints: nonpeptide beta-turn mimics. *Journal of Molecular Recognition*, 3 (2), 55–64.

[205] PATANI G & LAVOIE E (**1996**). Bioisosterism: A Rational Approach in Drug Design. *Chemical Reviews*, 96 (8), 3147–3176.

[206] DAVIDSON JP, LUBMAN O, ROSE T, WAKSMAN G, & MARTIN SF (**2002**). Calorimetric and structural studies of 1,2,3-trisubstituted cyclopropanes as conformationally constrained peptide inhibitors of Src SH2 domain binding. *Journal of the American Chemical Society*, 124 (2), 205–215.

[207] CORNISH-BOWDEN A (**2002**). Enthalpy-entropy compensation: a phantom phenomenon. *Journal of Biosciences*, 27 (2), 121–126.

[208] BEASLEY J, DOYLE D, CHEN L, COHEN D, FINE B, & PIELAK G (**2002**). Searching for quantitative entropy-enthalpy compensation among protein variants. *Proteins*, 49 (3), 398–402.

[209] LUNG FDT, LONG YQ, KING C, VARADY J, WU XW, WANG S, & ROLLER P (**2001**). Functional preference of the constituent amino acid residues in a phage-library-based non-phosphorylated inhibitor of the Grb2-SH2 domain. *Journal of Peptide Research*, 57, 447–454.

[210] OISHI S, KARKI R, SHI Z, WORTHY K, BINDU L, CHERTOV O, ESPOSITO D, FRANK P, GILLETTE W, MADERIA M, HARTLEY J, NICKLAUS M, BARCHI J JR, FISHER R, & BURKE T JR (**2005**). Evaluation of macrocyclic Grb2 SH2 domain-binding peptide mimetics prepared by ring-closing metathesis of C-terminal allylglycines with an N-terminal beta-vinyl-substituted phosphotyrosyl mimetic. *Bioorganic Medicinal Chemistry*, 13 (7), 2431–2438.

[211] NAM N, YE G, SUN G, & PARANG K (**2004**). Conformationally constrained peptide analogues of pTyr-Glu-Glu-Ile as inhibitiors of the Src SH2 domain binding. *Journal of Medicinal Chemistry*, 47, 3131–3141.

[212] DAVIDSON J & MARTIN S (**2000**). Use of 1,2,3-trisubstituted cyclopropanes as conformationally constained peptide mimics in SH2 antagonists. *Tetrahedron Letters*, 41, 9459–9464.

[213] BUCHANAN J, BOHACEK R, LUKE G, HATADA M, LU X, DALGARNO D, NARULA S, YUAN R, & DA H (**1999**). Structure-based design and synthesis of a novel class of Src SH2 inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 9, 2353–2358.

[214] BUCHANAN J, VU C, MERRY T, CORPUZ E, PRADEEPAN S, MANI U, YANG M, PLAKE H, VARKHEDKAR V, LYNCH B, MACNEIL I, LOIACONO K, TIONG C, & HOLT D (**1999**). Structure-activity relationships of a novel class of Src SH2 inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 9, 2359–2364.

[215] DEPREZ A, MANDINE E, VERMOND A, & LESUISSE D (**2002**). Imidazole-based ligands of the Src SH2 protein. *Bioorganic and Medicinal Chemistry Letters*, 12, 1287–1289.

[216] SHAKESPEARE WC, BOHACEK RS, AZIMIOARA MD, MACEK KJ, LUKE GP, DALGARNO DC, HATADA MH, LU XD, VIOLETTE SM, BARTLETT C, & SAWYER TK (**2000**). Structure-based design of novel bicyclic nonpeptide inhibitors for the Src SH2 domain. *Journal of Medicinal Chemistry*, 43 (21), 3815–3819.

[217] LESUISSE D, DEPREZ P, ALBERT E, DUC T, SORTAIS B, GOFFLO D, JEAN-BAPTISTE V, MARQUETTE JP, SCHOOT B, SARUBBI E, LANGE G, BROTO P, & MANDINE E (**2001**). Discovery of Thioazepinone Ligands for Src SH2: From Non-specific to Specific Binding. *Bioorganic and Medicinal Chemistry Letters*, 11 (16), 2127–2131.

[218] DEPREZ P, BAHOLET I, BURLET A, LANGE G, AMENGUAL R, SCHOOT B, VERMON A, MANDINE E, & LESUISSE D (**2002**). Discovery of highly potent Src SH2 binders: Structure-activity studies and X-ray structures. *Bioorganic and Medicinal Chemistry Letters*, 12, 1291–1294.

[219] HENRIQUES DA, LADBURY JE, & JACKSON RM (**2000**). Comparison of binding energies of SrcSH2-phosphotyrosyl peptides with structure-based prediction using surface area based empirical parameterization. *Protein Science*, 9, 1975–1985.

[220] PLUMMER MS, HOLLAND DR, SHAHRIPOUR A, LUNNEY EA, FERGUS JH, MARKS JS, MCCONNELL P, MUELLER WT, & SAWYER TK (**1997**). Design, synthesis, and cocrystal structure of a nonpeptide Src SH2 domain ligand. *Journal of Medicinal Chemistry*, 40 (23), 3719–3725.

[221] PACOFSKY GJ, LACKEY K, ALLIGOOD KJ, BERMAN J, CHARIFSON PS, CROSBY RM, DORSEY GF, FELDMAN PL, GILMER TM, HUMMEL CW, JORDAN SR, MOHR C, SHEWCHUK LM, STERNBACH DD, & RODRIGUEZ M (**1998**). Potent dipeptide inhibitors of the pp60(c-src) SH2 domain. *Journal of Medicinal Chemistry*, 41 (11), 1894–1908.

[222] METCALF CA, EYERMANN CJ, BOHACEK RS, HARALDSON CA, VARKHEDKAR VM, LYNCH BA, BARTLETT C, VIOLETTE SM, & SAWYER TK (**2000**). Structure-based design and solid-phase parallel synthesis of phosphorylated nonpeptides to explore hydrophobic binding at the Src SH2 domain. *Journal of Combinatorial Chemistry*, 2 (4), 305–313.

[223] LADBURY JE, LEMMON MA, ZHOU M, GREEN J, BOTFIELD MC, & SCHLESSINGER J (**1995**). Measurement of the Binding of Tyrosyl Phosphopeptides to SH2 Domains - a Reappraisal. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (8), 3199–3203.

[224] SCHAGGER H & VON JAGOW G (**1987**). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry*, 166, 368–379.

[225] DELAGLIO F, GRZESIEK S, VUISTER GW, ZHU G, PFEIFER J, & BAX A (**1995**). NMRPIPE - a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6, 277–293.

[226] KRAULIS P, DOMAILLE P, CAMPBELL-BURK S, VAN AKEN T, & LAUE E (**1994**). Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry*, 33 (12), 3515–3531.

[227] WISEMAN T, WILLISTON S, BRANDTS J, & LIN LN (**1989**). Rapid measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Analytical Biochemistry*, 179, 131–137.

[228] LADBURY J & CHOWDHRY B (editors) (**1998**). *Biocalorimetry. Applications of calorimetry in the biological sciences.* Wiley.

[229] SIGURSKJOLD B, BERLAND C, & SVENSSON B (**1994**). Thermodynamics of inhibitor binding to the catalytic site of glucoamylase from Aspergillus niger determined by displacement titration calorimetry. *Biochemistry*, 33 (33), 10191–10199.

[230] ZHANG YL & ZHANG ZY (**1998**). Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: A thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. *Analytical Biochemistry*, 261 (2), 139–148.

[231] BERMAN HM, WESTBROOK J, FENG Z, GILLILAND G, BHAT TN, WEISSIG H, SHINDYALOV IN, & BOURNE PE (**2000**). The Protein Data Bank. *Nucleic Acids Research*, 28, 235–242.

[232] HOLLAND D, LUNNEY E, PLUMMER M, MUELLER W, MCCONNELL P, PAVLOVSKY A, PARA K, SHAHRIPOUR A, HUMBLET C, SAWYER T, & RUBIN J (**1997**). Novel Pp60Src Sh2 Domain Crystal Structures: A 2.0 Angstrom Co-Crystal Structure of a D-Amino Acid Substituted Phosphopeptide Complex and a 2.1 Angstrom Apo Structure Displaying Self-Association. Unpublished report associated with 1BKL entry of PDB.

[233] METZLER WJ, LEITING B, PRYOR K, MUELLER L, & FARMER BT II (**1996**). The Three-Dimensional Solution Structure of the SH2 Domain from $p55^{blk}$ Kinase. *Biochemistry*, 35 (20), 6201 –6211.

[234] AROLD ST, ULMER TS, MULHERN TD, WERNER JM, LADBURY JE, CAMPBELL ID, & NOBLE MEM (**2001**). The role of the Src homology 3-Src homology 2 interface in the regulation of Src kinases. *Journal of Biological Chemistry*, 276 (20), 17199–17205.

[235] ZHANG W, SMITHGALL T, & GMEINER W (**1997**). Sequential assignment and secondary structure determination for the Src homology 2 domain of hematopoietic cellular kinase. *FEBS Letters*, 406 (1-2), 131–135.

[236] HABECK M, RIEPING W, LINGE J, & NILGES M (**2004**). NOE assignment with ARIA 2.0: the nuts and bolts. *Methods Molecular Biology*, 278, 379–402.

[237] WISHART D & SYKES B (**1994**). The 13C chemical shift index. A simple method for the identification of protein secondary structure using 13C chemical shift data. *Journal of Biomolecular NMR*, 4, 171–180.

[238] KARPLUS M (**1963**). Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society*, 85 (18), 2870–2871.

[239] VUISTER GW & BAX A (**1993**). Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J($H^N H\alpha$) Coupling Constants in ^{15}N-Enriched Proteins. *Journal of the American Chemical Society*, 115, 7772–7777.

[240] WILLIAMS M & LADBURY J (**2003**). *"Hydrogen bonds in protein-ligand complexes", in 'Molecular Recognition in Protein-Ligand Interactions'.* John Wiley-VCH, New York.

[241] MCDONALD IK & THORNTON JM (**1994**). Satisfying Hydrogen Bonding Potential in Proteins. *Journal of Molecular Biology*, 238, 777–793.

[242] GRZESIEK S, CORDIER F, & DINGLEY A (**2001**). Scalar couplings across hydrogen bonds. *Methods in Enzymology*, 338, 111–133.

[243] BAKER EN & HUBBARD RE (**1984**). Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology*, 44, 97–197.

[244] STICKLE DF, PRESTA LG, DILL KA, & ROSE GD (**1992**). Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Journal of Molecular Biology*, 226, 1143–1159.

[245] THORNTON JM, MACARTHUR MW, MCDONALD IK, JONES DT, MITCHELL JBO, NANDI CL, PRICE SL, & ZVELEBIL JJM (**1993**). Protein structures and complexes: what they reveal about the interactions which stabilise them. *Philosophical Transactions of The Royal Society of London. Series A, Mathematical and Physical Sciences*, 345, 113–129.

[246] MUMENTHALER C, GÜNTERT P, BRAUN W, & WÜTHRICH K (**1997**). Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *Journal of Biomolecular NMR*, 10 (4), 351–362.

[247] GÜNTERT P, MUMENTHALER C, & WÜTHRICH K (**1997**). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology*, 273 (1), 283–298.

[248] MOSELEY H & MONTELIONE G (**1999**). Automated analysis of NMR assignments and structures for proteins. *Current Opinion in Structural Biology*, 9 (5), 635–642.

[249] LINGE J, O'DONOGHUE S, & NILGES M (**2001**). Automated Assignment of Ambiguous Nuclear Overhauser Effects with ARIA. *Methods in Enzymology*, 339, 71–90.

[250] DUGGAN B, LEGGE G, DYSON H, & WRIGHT P (**2001**). SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. *Journal of Biomolecular NMR*, 19 (4), 321–329.

[251] GRONWALD W, MOUSSA S, ELSNER R, JUNG A, GANSLMEIER B, TRENNER J, KREMER W, NEIDIG K, & KALBITZER H (**2002**). Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). *Journal of Biomolecular NMR*, 23 (4), 271–287.

[252] HERRMANN T, GÜNTERT P, & WÜTHRICH K (**2002**). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*, 319 (1), 209–227.

[253] FOSSI M, LINGE J, LABUDDE D, LEITNER D, NILGES M, & H O (**2005**). Influence of chemical shift tolerances on NMR structure calculations using ARIA protocols for assigning NOE data. *Journal of Biomolecular NMR*, 31 (1), 21–34.

[254] LASKOWSKI RA, RULLMANN JAC, MACARTHUR MW, KAPTEIN R, & THORNTON JM (**1996**). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8, 477–486.

[255] KORADI R, BILLETER M, & WÜTHRICH K (**1996**). MOLMOL: a program fo display and analysis of macromolecular structures. *Journal of Molecular Graphics*, 14, 51–60.

[256] SINGER AU & FORMAN-KAY JD (**1997**). pH titration studies of an SH2 domain-phosphopeptide complex: unusual histidine and phosphate pKa values. *Protein Sci*, 6 (9), 1910–9.

[257] KIRKPATRICK S, GELATT C JR, & VECCHI M (**1983**). Optimization by Simulated Annealing. *Science*, 220 (4598), 671–680.

[258] LINGE J & NILGES M (**1999**). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *Journal of Biomolecular NMR*, 13 (1), 51–59.

[259] KONNERT J & HENDRICKSON W (**1980**). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallographica*, A36, 344–350.

[260] STEIN E, RICE L, & BRUNGER A (**1997**). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance*, 124 (1), 154–164.

[261] DORELEIJERS J, RAVES M, RULLMANN T, & KAPTEIN R (**1999**). Completeness of NOEs in protein structures: A statistical analysis of NMR data. *Journal of Biomolecular NMR*, 14, 123–132.

[262] HUMPHREY W, DALKE A, & SCHULTEN K (**1996**). VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14 (1), 33–38.

[263] JIMÉNEZ M, NIETO J, RICO M, SANTORO J, HERRANZ J, & BERMEJO F (**1986**). A study of the NH NMR signals of Gly-Gly-X-Ala tetrapeptides in H_2O at low-temperature. *Journal of Molecular Structure*, 143, 435–438.

[264] CORDIER F & GRZESIEK S (**1999**). Direct observation of hydrogen bonds in proteins by interresidue $^{3H}J_{NC}$ scalar couplings. *Journal of the American Chemical Society*, 121 (7), 1601–1602.

[265] DINGLEY A & GRZESIEK S (**1998**). Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide 2JNN couplings. *Journal of the American Chemical Society*, 120, 8293–8297.

[266] ARRINGTON CB & ROBERTSON AD (**2000**). Kinetics and Thermodynamics of Conformational Equilibria in Native Proteins by Hydrogen Exchange. *Methods in Enzymology*, 323, 104–127.

[267] CIERPICKI T & OTLEWSKI J (**2001**). Amide proton temperature coefficients as hydrogen bond indicators in proteins. *Journal of Biomolecular NMR*, 21, 249–261.

[268] PEDERSEN T, THOMSEN N, ANDERSEN K, MADSEN J, & POULSEN F (**1993**). Determination of the rate constants k1 and k2 of the Linderstrom-Lang model for protein amide hydrogen exchange. A study of the individual amides in hen egg-white lysozyme. *Journal of Molecular Biology*, 230 (2), 651–660.

[269] WOODWARD C, SIMON I, & TÜCHSEN E (**1982**). Hydrogen exchange and the dynamic structure of proteins. *Molecular and Cellular Biochemistry*, 48 (3), 135–160.

[270] BAXTER NJ & WILLIAMSON MP (**1997**). Temperature dependence of ^{1}H chemical shifts in proteins. *Journal of Biomolecular NMR*, 9, 359–369.

[271] FRAUENFELDER H, PETSKO G, & TSERNOGLOU D (**1979**). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature*, 280, 558–563.

[272] FRAUENFELDER H, HARTMANN H, KARPLUS M, KUNTZ I, KURIYAN J, PARAK F, PETSKO G, RINGE D, TILTON R JR, CONNOLLY M, & MAX N (**1987**). Themal Expansion of a Protein. *Biochemistry*, 26, 254–261.

[273] TILTON R JR, DEWAN J, & PETSKO G (**1992**). Effects of Temperature on Protein Structure and Dynamics: X-ray Crystallographic Studies of the Protein Ribonuclease-A at Nine Different Temperatures from 98 to 320 K. *Biochemistry*, 31, 2469–2481.

[274] KOPPLE K, OHNISHI M, & GO A (**1969**). Conformations of Cyclic Peptides. III. Cyclopentaglycyltyrosyl and Related Compounds. *Journal of the American Chemical Society*, 91, 4264.

[275] WAGNER G, PARDI A, & WÜTHRICH K (**1983**). Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *European Journal of Biochemistry*, 137 (3), 445–454.

[276] NICHOLLS A, SHARP K, & HONIG B (**1991**). GRASP. *Proteins: Structure Function and Genetics*, 11 (4), 281.

[277] BOHACEK RS, DALGARNO DC, HATADA M, JACOBSEN VA, LYNCH BA, MACEK KJ, MERRY T, METCALF CA, NARULA SS, SAWYER TK, SHAKESPEARE WC, VIOLETTE SM, & WEIGELE M (**2001**). X-ray structure of citrate bound to Src SH2 leads to a high-affinity, bone-targeted Src SH2 inhibitor. *Journal of Medicinal Chemistry*, 44 (5), 660–663.

[278] WEBER T, SCHAFFHAUSEN B, LIU YX, & GUNTHER UL (**2000**). NMR structure of the N-SH2 of the p85 subunit of phosphoinositide 3-kinase complexed to a doubly phosphorylated peptide reveals a second phosphotyrosine binding site. *Biochemistry*, 39 (51), 15860–15869.

[279] CICERO DO, MELINO S, ORSALE M, BRANCATO G, AMADEI A, FORLANI F, PAGANI S, & PACI M (**2003**). Structural rearrangements of the two domains of *Azotobacter vinelandii* rhodanese upon sulfane sulfur release: essential molecular dynamics, ^{15}N NMR relaxation and deuterium exchange on the uniformly labeled protein. *International Journal of Biological Macromolecules*, 33 (4-5), 193–201.

[280] WAKSMAN G (**1994**). Crystal-Structure of the Phosphotyrosine Recognition Domain SH2 of the Src Oncogene Product Complexed with Tyrosine-Phosphorylated Peptides. *Cellular and Molecular Biology*, 40 (5), 611–618.

[281] LIPARI G & SZABO A (**1982**). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *Journal of the American Chemical Society*, 104 (17), 4546–4559.

[282] CLORE GM, SZABO A, BAX A, KAY LE, DRISCOLL PC, & GRONENBORN AM (**1990**). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *Journal of the American Chemical Society*, 112 (12), 4989–4991.

[283] KEMPF J & LORIA J (**2003**). Protein dynamics from solution NMR. *Cell Biochemistry and Biophysics*, 37, 187–211.

[284] KAY L, TORCHIA D, & BAX A (**1989**). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28 (23), 8972–8979.

[285] KORZHNEV D, OREKHOV V, & ARSENIEV A (**1997**). Model-free approach beyond the borders of its applicability. *Journal of Magnetic Resonance*, 127 (2), 184–191.

[286] KRISTENSEN S, SIEGAL G, SANKAR A, & DRISCOLL P (**2000**). Backbone Dynamics of the C-terminal SH2 domain of the p85α subunit of Phosphoinositide 3-kinase: Effect of Phosphotyrosine-Peptide Binding and Characterization of Slow Conformational Exchange Processes. *Journal of Molecular Biology*, 299, 771–788.

[287] FINERTY PJ, MUHANDIRAM R, & FORMAN-KAY JD (**2002**). Side-chain dynamics of the SAP SH2 domain correlate with a binding hot spot and a region with conformational plasticity. *Journal of Molecular Biology*, 322 (3), 605–620.

[288] ZHANG W, SMITHGALL T, & GMEINER W (**1998**). Self-association and backbone dynamics of the hck SH2 domain in the free and phosphopeptide-complexed forms. *Biochemistry*, 37 (20), 7119–7126.

[289] PASCAL S, SINGER A, KAY L, & FORMANKAY J (**1995**). Refined Structure and Side-Chain Dynamics of an SH2 Domain from Plc-Gamma-1 Complexed with a pY1021 Peptide from Pdgfr. *Journal of Cellular Biochemistry*, pp. 33–33.

[290] VILLANUEVA J, HOSHINO M, KATOU H, KARDOS J, HASEGAWA K, NAIKI H, & GOTO Y (**2004**). Increase in the conformational flexibility of beta 2-microglobulin upon copper binding: a possible role for copper in dialysis-related amyloidosis. *Protein Science*, 13 (3), 797–809.

[291] BLUMENSCHEIN T, STONE D, FLETTERICK R, MENDELSON R, & SYKES B (**2005**). Calcium-dependent changes in the flexibility of the regulatory domain of TnC in the troponin complex. In press.

[292] PASCAL S, YAMAZAKI T, SINGER A, KAY L, & FORMAN-KAY J (**1995**). Structural and dynamic characterization of the phosphotyrosine binding region of a Src homology 2 domain–phosphopeptide complex by NMR relaxation, proton exchange, and chemical shift approaches. *Biochemistry*, 34 (36), 11353–11362.

[293] STRYER L (**1995**). *Biochemistry*. W.H. Freeman and Company.

[294] HARRIS T & TURNER G (**2002**). Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUMBM Life*, 53 (2), 85–98.

[295] MARENGERE L & PAWSON T (**1992**). Identification of residues in GTPase-activating protein Src homology 2 domains that control binding to tyrosine phosphorylated growth factor receptors and p62. *Journal of Biological Chemistry*, 267, 22779–22786.

[296] OLIVIER JP, RAABE T, HENKEMEYER M, DICKSON B, MBAMALU G, MARGOLIS B, SCHLESSINGER J, HAFEN E, & PAWSON T (**1993**). A Drosophila SH2-SH3 adaptor protein implicated in coupling the sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. *Cell*, 73 (1), 179–191.

[297] LEE CH, KOMINOS D, JACQUES S, MARGOLIS B, SCHLESSINGER J, SHOELSON SE, & KURIYAN J (**1994**). Crystal-Structures of Peptide Complexes of the Amino-Terminal SH2 Domain of the Syp Tyrosine Phosphatase. *Structure*, 2 (5), 423–438.

[298] BOERNER RJ, CONSLER TG, GAMPE RT, WEIGL D, WILLARD DH, DAVIS DG, EDISON AM, LOGANZO F, KASSEL DB, XU RX, PATEL IR, ROBBINS JS, LANSING T, GILMER TM, LUTHER MA, & KNIGHT WB (**1995**). Catalytic Activity of the SH2 Domain of Human pp60(c-Src), Evidence from NMR, Mass-Spectrometry, Site-Directed Mutagenesis and Kinetic-Studies for an Inherent Phosphatase-Activity. *Biochemistry*, 34 (46), 15351–15358.

[299] STUCKEY J, SCHUBERT H, FAUMAN E, ZHANG Z, DIXON J, & SAPER M (**1994**). Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 A and the complex with tungstate. *Nature*, 370 (6490), 571–575.

[300] KORTEMME T & CREIGHTON T (**1995**). Ionisation of Cysteine Residues at the Termini of Model α-Helical Peptides. Relevance to Unusual Thiol pK_a Values in Proteins of the Thioredoxin Family. *Journal of Molecular Biology*, 253, 799–812.

[301] PÉREZ-CAÑADILLAS J, CAMPOS-OLIVAS R, LACADENA J, DEL MARTINEZ P, GAVILANES J, SANTORO J, RICO M, & BRUIX M (**1998**). Characterization of pK_a values and titration shifts in the cytotoxic ribonuclease α-sarcin by NMR. Relationship between electrostatic interactions, structure, and catalytic function. *Biochemistry*, 37 (45), 15865–15876.

[302] JOSHI M, SIDHU G, NIELSEN J, BRAYER G, WITHERS S, & MCINTOSH L (**2001**). Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. *Biochemistry*, 40 (34), 10115–10139.

[303] TANFORD C (**1962**). The interpretation of hydrogen ion titration curves of proteins. *Advances in Protein Chemistry*, 17, 69–165.

[304] TANOKURA M (**1983**). ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine residues. I. Microscopic pK values and molar ratios of tautomers in histidine-containing peptides. *Biochimica et Biophysica Acta*, 742 (3), 576–585.

[305] MIEYAL J, STARKE D, GRAVINA S, & HOCEVAR B (**1991**). Thioltransferase in human red blood cells: kinetics and equilibrium. *Biochemistry*, 30 (36), 8883–8891.

[306] NELSON J & CREIGHTON T (**1994**). Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation *in vivo*. *Biochemistry*, 33 (19), 5974–5983.

[307] PETERS G, FRIMURER T, & OLSEN O (**1998**). Electrostatic Evaluation of the Signature Motif (H/V)CX_5R(S/T) in Protein-Tyrosine Phosphatases. *Biochemistry*, 37, 5383–5393.

[308] KRISHNA M, HOANG L, LIN Y, & ENGLANDER S (**2004**). Hydrogen exchange methods to study protein folding. *Methods*, 34 (1), 51–64.

[309] JONES G, WILLET P, & GLEN RC (**1995**). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, 245, 43–53.

[310] CLAUSSEN H, BUNING C, RAREY M, & LENGAUER T (**2001**). FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *Journal of Molecular Biology*, 308, 377–395.

[311] RAREY M, KRAMER B, LENGAUER T, & KLEBE G (**1996**). A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology*, 261, 470–489.

[312] KLEBE G & MIETZNER T (**1994**). A fast and efficient method to generate biologically relevant conformations. *Journal of Computer-Aided Molecular Design*, 8, 583–606.

[313] ALLEN FH (**2002**). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica*, B58, 380–388.

[314] DECLERCQ PJ (**1984**). Systematic conformational analysis. A microcomputer method for the semiquantitative evaluation of polycyclic systems containing five-, six-, and seven-membered rings. 1. Program characteristics. *Tetrahedron*, 40 (19), 3717–3727.

[315] BÖHM HJ (**1994**). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of a known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, 8, 243–256.

[316] TRIPOS Associates, Inc., St Louis, Missouri, USA. SYBYL *Molecular Modelling Software Version 6.x.*

[317] CONNOLLY ML (**1983**). Analytical molecular surface calculation. *Journal of Applied Crystallography*, 16, 548–558.

[318] TANIMOTO T (**1957**). IBM Internal Report.

[319] BROWN R & MARTIN Y (**1996**). Use of Structure-Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *Journal of Chemical Information and Computer Sciences*, 36 (3), 572–584.

[320] PEARLMAN RS & SMITH KM (**1998**). Novel software tools for chemical diversity. *Perspectives in Drug Discovery and Design*, 9/10/11, 339–353.

[321] POY F, YAFFE M, SAYOS J, SAXENA K, MORRA M, SUMEGI J, CANTLEY L, TERHORST C, & ECK M (**1999**). Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Molecular Cell*, 4 (4), 555–561.

[322] LI SC, GISH G, YANG DW, COFFEY AJ, FORMAN-KAY JD, ERNBERG I, KAY LE, & PAWSON T (**1999**). Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A. *Current Biology*, 9 (23), 1355–1362.

[323] YE B, AKAMATSU M, SHOELSON S, WOLF G, GIORGETTI-PERALDI S, YAN X, ROLLER P, & BURKE T JR (**1995**). L-O-(2-malonyl)tyrosine: a new phosphotyrosyl mimetic for the preparation of Src homology 2 domain inhibitory peptides. *Journal of Medicinal Chemistry*, 38, 4270–4275.

[324] PEARLMAN R (**1987**). Rapid Generation of High Quality Approximate 3D Molecular Structures. *Chemical Design and Automation News*, 2 (1), 5–7.

[325] MENG C, SHOICHET BK, & KUNTZ ID (**1992**). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, 13 (4), 505–524.

[326] ELDRIDGE MD, MURRAY CW, AUTON TR, PAOLININE GV, & MEE RP (**1997**). Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-aided Molecular Design*, 11, 425–445.

[327] MUEGGE I & MARTIN YC (**1999**). A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *Journal of Medicinal Chemistry*, 42, 791–804.

[328] GASTEIGER J & MARSILI M (**1980**). Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron*, 36 (22), 3219–3228.

[329] HAHN EL (**1950**). Spin Echoes. *Physical Review*, 80 (4), 580–594.

[330] CLARIDGE TDW (**1999**). *High Resolution NMR Techniques in Organic Chemistry*, volume 19 of *Tetrahedron Organic Chemistry Series*. Pergamon.

[331] CARR H & PURCELL E (**1954**). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review*, 94 (3), 630–638.

[332] PRICE N & DWEK R (**1996**). *Principles and Problems in Physical Chemistry for Biochemists*. Oxford Science Publications.

[333] VOLD R, WAUGH J, KLEIN M, & PHELPS D (**1968**). Measurement of Spin Relaxation in Complex Systems. *Journal of Chemical Physics*, 48 (8), 3831–3832.

[334] MERCIER K & POWERS R (**2005**). Determining the optimal size of small molecule mixtures for high throughput NMR screening. *Journal of Biomolecular NMR*, 31 (3), 243–258.

[335] COSTANTINO L & BARLOCCO D (**2004**). Inhibitors for proteins endowed with catalytic and non-catalytic activity which recognize pTyr. *Current Medicinal Chemistry*, 11 (20), 2725–2747.

[336] SCHULTZ C (**2003**). Prodrugs of Biologically Active Phosphate Esters. *Bioorganic and Medicinal Chemistry*, 11, 885–898.

[337] LIU WQ, VIDAL M, OLSZOWY C, MILLION E, LENOIR C, DHOTEL H, & GARBAY C (**2004**). Structure-activity relationships of small phosphopeptides, inhibitors of Grb2 SH2 domain, and their prodrugs. *Journal of Medicinal Chemistry*, 47 (5), 1223–1233.

[338] HOFBAUER C & ASZODI A (**2005**). SH2 binding site comparison: a new application of the SURFCOMP method. *Journal of Chemical Information and Modeling*, 45 (2), 414–421.

[339] HUANG N, NAGARSEKAR A, XIA G, HAYASHI J, & MACKERELL A JR (**2004**). Identification of non-phosphate-containing small molecular weight inhibitors of the tyrosine kinase p56 Lck SH2 domain via in silico screening against the pY + 3 binding site. *Journal of Medicinal Chemistry*, 47 (14), 3502–3511.

[340] MAINA F, CASAGRANDA F, AUDERO E, SIMEONE A, COMOGLIO P, KLEIN R, & PONZETTO C (**1996**). Uncoupling of Grb2 from the Met receptor in vivo reveals complex roles in muscle development. *Cell*, 87 (3), 531–542.

[341] McNemar C, Snow ME, Windsor WT, Prongay A, Mui P, Zhang RM, Durkin J, Le HV, & Weber PC (**1997**). Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry*, 36 (33), 10006–10014.

[342] Rogers J, Cheng H, & Smithgall T (**2000**). Src homology 2 domain substitution modulates the kinase and transforming activities of the Fes protein-tyrosine kinase. *Cell Growth and Differentiation*, 11, 581–592.

[343] Gee K, Sun W, Bhalgat M, Upson R, Klaubert D, Latham K, & Haugland R (**1999**). Fluorogenic substrates based on fluorinated umbelliferones for continuous assays of phosphatases and beta-galactosidases. *Analytical Biochemistry*, 273 (1), 41–48.

[344] Venetz W, Mangan C, & Siddiqi I (**1990**). Kinetic determination of alkaline phosphatase activity based on hydrolytic cleavage of the P-F bond in monofluorophosphate and fluoride ion-selective electrode. *Analytical Biochemistry*, 191 (1), 127–132.

[345] Kuerová Z & Skoda J (**1971**). Resistance of uridine 5'-fluorophosphate to alkaline phosphatase and its sensitivity to 5'-nucleotidase. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 247 (2), 194–196.

[346] Hertzberg R & Pope A (**2000**). High-throughput screening: new technology for the 21st century. *Current Opinion in Chemical Biology*, 4 (4), 445–451.

[347] Liu B, Li S, & Hu J (**2004**). Technological advances in high-throughput screening. *American Journal of Pharmacogenomics*, 4 (4), 263–276.

[348] Owicki JC (**2000**). Fluorescence Polarization and Anisotropy in High Throughput Screening: Perspectives and Primer. *Journal of Biomolecular Screening*, 5 (5), 297–306.

[349] Pastula C, Johnson I, Beechem J, & Patton W (**2003**). Development of fluorescence-based selective assays for serine/threonine and tyrosine phosphatases. *Combinatorial Chemistry and High Throughput Screening*, 6, 341–346.